



IMPERIAL AGRICULTURAL  
RESEARCH INSTITUTE, NEW DELHI.

19334

MGIPC—84—III-1-93—22-8-45—5,000.







**PROCEEDINGS**  
OF THE  
**NATIONAL ACADEMY**  
**OF SCIENCES**  
OF THE  
UNITED STATES OF AMERICA

**VOLUME 28, 1942**



**EXECUTIVE COMMITTEE**  
OF THE PROCEEDINGS

**EDWIN BIDWELL WILSON, *Managing Editor***

**R. A. MILLIKAN**

**R. G. HARRISON**

---

***PUBLISHED MONTHLY***

**Publication Office: Mack Printing Company, Easton, Pa.**

**Editorial Office: Harvard School of Public Health, Boston 17**

**Home Office of the Academy: Washington, D. C.**



# CONTENTS

## MATHEMATICS

	Page
REPRESENTATION OF MEASURABLE FLOWS IN EUCLIDEAN 3-SPACE.....	
<i>By S. Kakutani</i>	16
INVERSION OF A GENERALIZED LAPLACE INTEGRAL.....	
<i>By R. P. Boas, Jr.</i>	21
ON FORMAL EXPONENTIAL DIFFERENTIATION IN RINGS.....	
<i>By H. S. Vandiver</i>	24
A GENERAL THEOREM ON THE INITIAL CURVATURE OF DYNAMICAL TRAJECTORIES	
<i>By E. Kasner and D. Mittleman</i>	48
GENERALIZED TRANSFORMATION THEORY OF ISOTHERMAL AND DUAL FAMILIES...	
<i>By E. Kasner and J. De Cicco</i>	52
CONCERNING SEPARABILITY.....	
<i>By R. L. Moore</i>	56
THE COEFFICIENTS OF STIRLING'S SERIES FOR $\log \Gamma(x)$ .....	
<i>By H. S. Uhler</i>	59
OSCILLATIONS OF THE DERIVATIVES OF A FUNCTION.....	
<i>By A. C. Schaeffer</i>	62
MAXIMAL SYLOW SUBGROUPS OF A GIVEN GROUP.....	
<i>By G. A. Miller</i>	80
AN INVERSE PROBLEM CONCERNING A CHAIN PROCESS.....	
<i>By I. Opatowski</i>	83
ON CONFIDENCE INTERVALS.....	
<i>By E. B. Wilson</i>	88
ON CONTINGENCY TABLES.....	
<i>By E. B. Wilson</i>	94
CONCERNING THE DIFFERENTIAL EQUATIONS OF SOME BOUNDARY LAYER PROBLEMS. II.....	
<i>By H. Weyl</i>	100
UNSTABLE MINIMAL SURFACES WITH ANY RECTIFIABLE BOUNDARY.....	
<i>By M. Shiffman</i>	103
ON HOMOGENEOUS MEASURE ALGEBRAS.....	
<i>By D. Maharam</i>	108
CERTAIN DIRECT PRODUCTS OF THE GROUPS OF SELF-ISOMETRIES.....	
<i>By G. A. Miller</i>	141
CERTAIN CONGRUENCE CRITERIA CONNECTED WITH FERMAT'S LAST THEOREM...	
<i>By H. S. Vandiver</i>	144
ON GENERALIZED BIORTHOGONAL EXPANSIONS IN METRIC AND UNITARY SPACES.	
<i>By Y. Y. Tseng</i>	170
REPRESENTATION OF ONE-PARAMETER SEMI-GROUPS OF LINEAR TRANSFORMATIONS.....	
<i>By E. Hille</i>	175
ON A CLASS OF SURFACES.....	
<i>By G. Fubini</i>	178
SOME DEDUCTIONS FROM FROBENIUS'S THEOREM.....	
<i>By G. A. Miller</i>	251
ON MONOTHETIC GROUPS.....	
<i>By P. R. Halmos and H. Samelson</i>	254
ON THE DIRICHLET PROBLEM FOR THE HYPERBOLIC CASE.....	
<i>By J. Hadamard</i>	258
STATIONARY POINTS OF TRANSFORMATION GROUPS.....	
<i>By P. A. Smith</i>	293
NOTE ON THE $t$ -TEST.....	
<i>By E. B. Wilson and J. Worcester</i>	297
GENERAL CONGRUENCES INVOLVING THE BERNOULLI NUMBERS.....	
<i>By H. S. Vandiver</i>	324
TRANSFORMATION THEORY OF ISOGONAL TRAJECTORIES OF ISOTHERMAL FAMILIES	
<i>By E. Kasner and J. De Cicco</i>	328
DIFFERENTIAL EQUATIONS OF THE TYPE: $y''' = Gy'' + Hy'^2$ .....	
<i>By E. Kasner</i>	333
AUTOMORPHISMS OF THE DIHEDRAL GROUPS.....	
<i>By G. A. Miller</i>	368
SOME ASYMPTOTIC RELATIONS.....	
<i>By H. Bateman</i>	371
AN ORTHOGONAL PROPERTY OF THE HYPERGEOMETRIC POLYNOMIAL.....	
<i>By H. Bateman</i>	374
CONTINGENCY TABLES.....	
<i>By E. B. Wilson and J. Worcester</i>	378
THE ASSOCIATION OF THREE ATTRIBUTES.....	
<i>By E. B. Wilson and J. Worcester</i>	384
THE PERMUTATION GROUPS OF A GENERAL DEGREE.....	
<i>By G. A. Miller</i>	407
THE STRUCTURE OF ALGEBRAIC MODULS.....	
<i>By G. Harrison</i>	410
SYNTHETIC SOLUTION OF THE INVERSE PROBLEM OF DYNAMICS.....	
<i>By E. Kasner and J. De Cicco</i>	413

ON THE THEORY OF ANALYTIC CURVES.....	By H. and J. Weyl	417
ON THE ANALYTICAL THEORY OF SEMI-GROUPS.....	By E. Hille	421
GENERALIZATION OF SUNDMAN'S FUNDAMENTAL EQUALITY TO THE CASE OF MORE THAN THREE BODIES.....	By G. Garcia	425
GENERALIZATION OF THE INEQUALITY OF SUNDMAN TO THE CASE OF MORE THAN THREE BODIES AND TO THE CASE OF A GRAVITATIONAL GAS....	By G. Garcia	428
STATISTICAL METRICS.....	By K. Menger	535
NATURAL ISOMORPHISMS IN GROUP THEORY.....	By S. Eilenberg and S. MacLane	537
CONCERNING INTERSECTING CONTINUA.....	By R. L. Moore	544
CONCERNING A CONTINUUM AND ITS BOUNDARY.....	By R. L. Moore	550
CONCERNING DOMAINS WHOSE BOUNDARIES ARE COMPACT.....	By R. L. Moore	555
ON SETS OF INTEGERS WHICH CONTAIN NO THREE TERMS IN ARITHMETICAL PROGRESSION.....	By R. Salem and D. C. Spencer	561
SOME ADVANCES IN THE COMBINATORY THEORY OF QUANTIFICATION.....	By H. B. Curry	564

### ASTRONOMY

THE CALCULATION OF ROTATION FACTORS FOR ECLIPSING BINARY SYSTEMS.....	By Z. Kopal	133
GALACTIC AND EXTRAGALACTIC STUDIES, XIII. NOTE ON THE COMPARATIVE DIAMETERS OF SPHEROIDAL AND SPIRAL GALAXIES.....	By H. Shapley	186
GALACTIC AND EXTRAGALACTIC STUDIES, XIV. ON THE MAGNITUDE DISPERSION IN THE PERIOD-LUMINOSITY RELATION.....	By H. Shapley, V. McKibben and R. A. Craig	192
GALACTIC AND EXTRAGALACTIC STUDIES, XV. ON THE DISTRIBUTION OF PERIODS FOR 343 CEPHEIDS IN THE SMALL MAGELLANIC CLOUD.....	By H. Shapley and V. McKibben	200
ON THE PERSEUS CLUSTER OF NEBULAE.....	By F. Zwicky	317
ON THE PHYSICAL CHARACTERISTICS OF THE PERSEUS CLUSTER OF NEBULAE....	By F. Zwicky	355
ON THE MATERIAL EJECTED FROM NOVAE.....	By C. Payne-Gaposchkin and S. Gaposchkin	482
ON THE DIMENSIONS AND CONSTITUTION OF VARIABLE STARS.....	By C. Payne-Gaposchkin and S. Gaposchkin	490
A STUDY OF THE RV TAURI VARIABLES.....	By C. Payne-Gaposchkin and V. K. Brenton	496
GALACTIC AND EXTRAGALACTIC STUDIES, XVI. PHOTOGRAPHIC AMPLITUDES OF CLASSICAL CEPHEIDS.....	By H. Shapley	501

### PHYSICS

ON THE DEPOLARIZATION OF NEUTRON BEAMS BY MAGNETIC FIELDS.....	By O. Halpern and T. Holstein	112
ON THE PHYSICAL CHARACTERISTICS OF THE HYDRA CLUSTER OF NEBULAE.....	By F. Zwicky	150
THE ENERGY EQUATION FOR A VISCOUS COMPRESSIBLE FLUID... By H. J. Stewart		161
THE DISTRIBUTION OF METALLIC ATOMS IN TWO-COMPONENT GLASSES.....	By J. S. Lukash	277
A NEW FOURIER SERIES TECHNIQUE FOR CRYSTAL STRUCTURE DETERMINATION.	By M. J. Buerger	281

## CHEMISTRY

RADIOACTIVE CARBON AS AN INDICATOR OF CARBON DIOXIDE UTILIZATION. VIII.	
THE RÔLE OF CARBON DIOXIDE IN CELLULAR METABOLISM.....	
By C. B. van Niel, S. Ruben, S. F. Carson, M. D. Kamen and J. W. Foster	8
THE OXIDATION OF 3,5-DIHYDROXYTYROSINE TO THYROXINE.....	
By T. B. Johnson and L. B. Tewkesbury	73
COMPLEMENT FIXATION WITH SIMPLE SUBSTANCES CONTAINING TWO OR MORE	
HAPTENIC GROUPS.....By D. Pressman, D. H. Campbell and L. Pauling	77
RADIOACTIVE CARBON AS AN INDICATOR OF CARBON DIOXIDE UTILIZATION. IX.	
THE ASSIMILATION OF CARBON DIOXIDE BY PROTOZOA.....	
By C. B. van Niel, J. O. Thomas, S. Ruben and M. D. Kamen	157
THE DISSOCIATION CONSTANT IN NITROGEN FIXATION BY <i>Azotobacter</i> .....	
By P. W. Wilson, R. H. Burris and C. J. Lind	243
PURINES AS GROWTH REQUIREMENTS OF <i>Spirillum Serpens</i> ..By D. E. Pennington	
	272
THE LOWER HYDRATES OF SOAP.....	
By M. J. Buerger, L. B. Smith, A. de Bretteville and F. V. Ryer	528
THE CHARACTERISTICS OF SOAP HEMIHYDRATE CRYSTALS.....By M. J. Buerger	
	529

## GEOLOGY AND PALEONTOLOGY

DISTORTION OF STRATIGRAPHIC THICKNESSES DUE TO FOLDING.....By E. Cloos	401
--	-----

## BOTANY

GUANINE AND FACTOR Z <sub>1</sub> , GROWTH SUBSTANCES FOR <i>Phycomyces</i> .....	
By W. J. Robbins and F. Kavanagh	4
COMPARATIVE RATES OF DIVISION IN LARGE AND SMALL CELLS OF DEVELOPING	
FRUITS.....By E. W. Sinnott	36
HYPOXANTHINE, A GROWTH SUBSTANCE FOR PHYCOMYCES.....	
By W. J. Robbins and F. Kavanagh	65
CONCENTRATION OF GROWTH HORMONE AND FRUITFULNESS IN THE MONTMORENCY	
CHERRY.....By F. G. Gustafson	131
THE RELEASE OF AUXIN FROM ISOLATED LEAF PROTEINS OF SPINACH BY ENZYMES	
By S. G. Wildman and S. A. Gordon	217
SYNTHESIS OF VITAMINS BY INTESTINAL BACTERIA.....	
By P. R. Burkholder and I. McVeigh	285
A REVERSIBLE GROWTH INHIBITION OF ISOLATED TOMATO ROOTS...By J. Bonner	
	321
A PYRIMIDINE ANALOG OF THIAMINE AND THE GROWTH OF FUNGI.....	
By W. J. Robbins	352
THE INCREASE OF B VITAMINS IN GERMINATING SEEDS.....	
By P. R. Burkholder and I. McVeigh	440
NOTES ON ALGAL NOMENCLATURE. I. <i>Pollexfenia</i> , <i>Jeannerettia</i> AND <i>Mesotrema</i>	
By G. F. Papenfuss	446
SEXUAL HORMONES IN <i>ACHLYA</i> . V. HORMONE A', A MALE-SECRETED AUG-	
MENTER OR ACTIVATOR OF HORMONE A.....By J. R. Raper	509

## ZOOLOGY AND ANATOMY

RELATIVE EFFECTS OF X-RAYS AND NEUTRONS ON CHROMOSOMES IN DIFFERENT	
PARTS OF THE "RESTING STATE".....By A. Marshak	29
SENSITIZATION OF MELANOPHORES BY NERVE CUTTING.....By G. H. Parker	
	164

THE EFFECT OF X-RAY STIMULATION ON THE BIOELECTRIC POTENTIALS OF THE AVIAN EGG.....	By A. L. Romanoff and A. A. Bless	306
NUCLEIC ACID STORAGE IN THE TOAD'S EGG.....	By T. S. Painter and A. N. Taylor	311
OXIDATION-REDUCTION PATTERNS IN AMPHIBIAN AND TELEOST DEVELOPMENT..	By C. M. Child	339
NUCLEOPROTEINS OF CELL NUCLEI.....	By A. E. Mirsky and A. W. Pollister	344
A COMPLEMENT-RELEASE REACTION; THE NEUTRALIZATION OF THE ANTICOMPLEMENTARY ACTION OF SEA-URCHIN FERTILIZIN BY ANTIFERTILIZIN.....	By A. Tyler	391
WHITE PIGMENTARY EFFECTORS ( <i>Leucophores</i> ) IN KILLIFISHES..	By E. F. B. Fries	396

## GENETICS

CHROMOSOME DEGENERATION IN RELATION TO GROWTH AND HYBRID VIGOR.....	By D. F. Jones	38
TEMPERATURE AND "SEX-RATIO" IN <i>Drosophila Pseudoöbscura</i> .....	By C. D. Darlington and T. Dobzhansky	45
SIZE GENES OF MICE.....	By W. E. Castle	69
AN AUTOSOMAL RECESSIVE FACTOR INDUCING SEMI-STERILITY IN <i>Drosophila Melanogaster</i> FEMALES.....	By U. Fano	119
"ATAXIA," A HEREDITARY NERVOUS DISORDER OF THE RABBIT.....	By P. B. Sawin, M. V. Anders and R. B. Johnson	123
CLOSE LINKAGE BETWEEN MUTATIONS WITH SIMILAR EFFECTS.....	By L. C. Dunn and E. Caspari	205
THE EFFECT OF pH ON INACTIVATION OF TOBACCO MOSAIC VIRUS BY X-RAYS... ..	By A. Marshak and W. N. Takahashi	211
THE DISTRIBUTION OF X-RAY INDUCED CHROMOSOMAL ABERRATIONS..	By K. Sax	229
GENETIC CONTROL OF BIOCHEMICAL REACTIONS IN NEUROSPORA: AN "AMINO-BENZOICLESS" MUTANT.....	By E. L. Tatum and G. W. Beadle	234
CHROMOSOMES OF THE RED FOX.....	By L. Wipf and R. M. Shackelford	265
SALIVARY GLAND TYPE CHROMOSOMES IN MOSQUITOES.....	By E. Sutton	268
DIFFUSION OF GENE PRODUCTS.....	By K. Sax	303
ON THE ANAPHASE MOVEMENT OF CHROMOSOMES.....	By M. M. Rhoades and H. Vilkomerson	433
HEAT-INDUCED TRIPLOIDY IN THE NEWT, <i>Triturus Viridescens</i> .....	By G. Fankhauser and R. C. Watson	436
THE FUSION OF BROKEN ENDS OF CHROMOSOMES FOLLOWING NUCLEAR FUSION..	By B. McClintock	458
NON-RANDOM UNCOILING IN HETEROBRACHIAL CHROMOSOMES..	By A. H. Sparrow	463
THE GENETIC NATURE OF X-RAY INDUCED CHANGES IN POLLEN..	By C. M. Rick	518

## PHYSIOLOGY AND BIOCHEMISTRY

THE RÔLE OF INTRACELLULAR CATIONS ON LIVER GLYCOGEN FORMATION <i>in Vitro</i>	By A. B. Hastings and J. M. Buchanan	478
---	--------------------------------------	-----

## PATHOLOGY AND BACTERIOLOGY

THE IDENTIFICATION AND CHARACTERIZATION OF BACTERIOPHAGES WITH THE ELECTRON MICROSCOPE.....	By S. E. Luria and T. F. Anderson	127
THE EPIDEMIC CURVE.....	By E. B. Wilson and M. H. Burke	361

SOME FACTORS AFFECTING THE TOXICITY OF CULTURES OF *Shigella dysenteriae*. . .  
By R. J. Dubos, H. D. Hoberman and C. Pierce 453

TUMOR FORMATION BY ATTENUATED CROWN-GALL BACTERIA IN THE PRESENCE  
OF GROWTH-PROMOTING SUBSTANCES. . . . . By A. C. Braun and T. Laskaris 468

OXIDIZED COTTON, AN IMMUNOLOGICALLY SPECIFIC POLYSACCHARIDE. . . . .  
By M. Heidelberger and G. M. Hobby 516

ANTHROPOLOGY AND PSYCHOLOGY

PREHISTORIC CONTACT BETWEEN SOUTH AMERICA AND THE WEST INDIES. . . . .  
By C. Osgood 1

THEORY OF THE EFFECTS OF LIGHT INTENSITY AND DURATION IN DETERMINING  
VISUAL RESPONSES. . . . . By N. R. Bartlett and G. F. Hudson 289





PROCEEDINGS  
OF THE  
NATIONAL ACADEMY OF SCIENCES

Volume 28

January 15, 1942

Number 1

*Copyright 1942 by the National Academy of Sciences*

---

*ANTHROPOLOGY.—PREHISTORIC CONTACT BETWEEN SOUTH  
AMERICA AND THE WEST INDIES*

BY CORNELIUS OSGOOD

DEPARTMENT OF ANTHROPOLOGY, YALE UNIVERSITY

Communicated December 3, 1941

The primary historical problem which has intrigued the archeologists of the Caribbean area has been that of the cultural relations between the West Indies and the two adjacent continents. A cultural connection with South America has for long been a generally accepted hypothesis, and more recently evidence has been increasing in support of the view that there were prehistoric contacts between the Greater Antilles and the United States<sup>1</sup> and Mexico.<sup>2</sup> In one case at least, hypothesis can now be set aside, for as a result of archeological work carried on during the past summer (1940), it is possible for the first time to link the West Indies with the South American mainland in an empirical manner which places the question of prehistoric contact beyond reasonable doubt. It is the purpose of this paper to indicate the nature of the evidence in support of this statement, and in a general way to correlate the newly discovered cognate mainland culture with the pattern of cultures previously recognized among the islands of the Caribbean.

A positive and provocative indication of direct contact appeared during the summer of 1933 when the writer, engaged in the Caribbean Archeological Program of the Yale Peabody Museum, found in the lowest stratum of a mound in the Lake Valencia region of Venezuela, a number of fragments of fine-grained, well-polished pottery with broadly incised lines and the so-called "doughnut eye." These specimens were absolutely atypical of the local culture and were described as trade sherds with West Indian affinities.<sup>3</sup>

In 1941, when the writer and Mr. George Howard of Yale University undertook an archeological survey of Venezuela, a culture was located on the lower Orinoco River which appeared to be the source of the trade sherds mentioned above.<sup>4</sup> Excavation in a site of this culture near the town of Barrancas produced fragmentary remains of a fine-grained pottery beautifully decorated with broadly incised curvilinear designs on lustrous

slips of red, black, yellow and brown. Outstanding among obvious features of the material were modeled *adornos* or ornamental heads of various animals which served as decorations for the sides of pots. A few days after the excavation at Barrancas was completed, the writer flew to Trinidad and had the pleasure of examining the material recently excavated by Mr. J. A. Bullbrook and Capt. J. E. L. Carter from the great shell midden at Erin Bay. This site is but a short distance across the Serpent's Mouth (as the intervening ocean channel is called) from the northern entrance to the Orinoco River, and only about 100 kilometers or 52 miles airline from the site at Barrancas. Some of the decorated pottery from the lower of the two strata at Erin Bay is to all appearances identical with that from the Orinoco River, and certainly such similarity exists as to rule out any likelihood of chance convergence in styles or even of sameness resulting only from casual contacts. As nearly as could be judged, numerous peculiar pieces from one site could be replaced by samples from the other without differences being readily detected. The similarity is further borne out by a comparison of the Barrancas material with that excavated at Erin Bay by Fewkes in the winter of 1912-1913 and illustrated in several of his papers.<sup>5</sup> Such is the present evidence for positive prehistoric contact between the West Indies and the South American mainland.

In order to correlate these connected cultures with the pattern of those recognized conditionally for the West Indies, we may first review the present theory regarding the population of the islands. From the linguistic and historical point of view, three major groups, the *Carib*, the *Arawak* and the *Ciboney*, appear to have occupied the Antilles in pre-Columbian times. These we shall consider in turn.

*Carib.* The most recent group in the West Indies is certainly the Carib, a people who are known historically to have raided and overrun most of the islands of the Lesser Antilles shortly before European contact.<sup>6</sup> Their South American origin is scarcely questioned but their culture, in terms of archeological remains, is very inadequately known. It may even be that any detailed knowledge of their culture is unobtainable, since the period of occupation by the Carib group possibly was too short to leave any accumulations sufficient for archeologists to interpret.

*Arawak.* The most widely spread of the major groups in the West Indies is the Arawak. Like the Carib, its presence is known from historic sources. Arawak occupation certainly antedated that of the Carib and is of larger duration, sufficiently long, indeed, so that it is not possible, on direct linguistic and historical evidence, to be certain that we are dealing with only a single group which, properly speaking, can be called Arawak. Rainey, working in Puerto Rico, described two cultures, an earlier Crab and a later Shell, which serve to distinguish phases of the Arawak group as I use the term here.<sup>7</sup> Although Rainey, with reasonable caution, did not

regard both cultures as Arawak, recent investigations by Rouse in Puerto Rico support this thesis.<sup>8</sup> In what is the most complete summary of West Indian archeology, Rainey went on to find widely spread evidences for his two cultures, identifying the Shell culture principally with the Greater Antilles and the Crab with the Lesser Antilles.<sup>9</sup> Regardless of revisions which may be made in his correlations, the existence of distinguishable earlier and later phases appears definitely established.

*Ciboney.* The third and most restricted of the major groups in the West Indies is the Ciboney which is best known from Cuba. Unlike the Carib and Arawak groups, the Ciboney people are not known from historical contacts but are referred to indirectly in the source literature. They clearly stand apart from the others, however, on archeological evidence, as they lacked pottery. Whenever they occupied the same territory as the Arawak (there is no evidence that the Ciboney and the protohistoric Carib ever shared the same islands), the Ciboney appear as the earlier group, which fact, however, by no means signifies that, in terms of actual time, they entered the West Indies before the Arawak. That the affinities of the Ciboney are with Florida is a widely proposed hypothesis<sup>10</sup> for which adequate empirical evidence, however, is yet to be established.

To return to the question of how the recently correlated sites of Barrancas, Venezuela and Erin Bay, Trinidad, fit into the general West Indian pattern, certain conclusions can be reached. From every point of view, the Erin Bay culture seems to belong to the Arawak group and it is the earlier phase of the remains which has direct ties with the Orinoco.<sup>11</sup> It is the feeling of some workers in Trinidad that this same phase correlates with the material at Cape Mayaro on the east coast of Trinidad which, we may note, has been described by Rainey as demonstrating a "remarkable similarity" with the Crab culture in Puerto Rico.<sup>12</sup> Thus it can be said that the Barrancas culture probably correlates with the early (Crab culture?) phase of the Arawak group in the West Indies.<sup>13</sup>

<sup>8</sup> Rouse, I., "Some Evidence Concerning the Origins of West Indian Pottery-Making," *Amer. Anth.*, 42, 49-80 (1940).

<sup>9</sup> Ries, M., "Summary Report on the Tulane University-Cuban Navy Expedition to Cabo San Antonio, Cuba," Manuscript (1937).

<sup>10</sup> Osgood, C., "Excavations Near Lake Valencia, Venezuela," Manuscript (1935).

<sup>11</sup> This work was financed through the Institute of Andean Research by the Committee for Cultural and Commercial Relations with the Latin American Republics. It is expected that detailed results will appear in the *Yale University Publications in Anthropology* during 1942-1943.

<sup>12</sup> Fewkes, J. W., "Prehistoric Objects from a Shell-Heap at Erin Bay, Trinidad," *Amer. Anth.*, 16, 200-220 (1914). Reprinted in *Contrib. from Heye Museum*, 1 (1915). The same illustrations are also available in Fewkes, J. W., "A Prehistoric Island Culture Area of America," *Bur. Amer. Ethnol.*, An. Rep. 34, 70, plates 2-8 (1922).

<sup>13</sup> Loven, one of the principal writers on the prehistory of the West Indies but one who depends primarily on the historical sources, holds that the *Caribs* were in Trinidad

as well as the other small islands of the Lesser Antilles. In this view Bullbrook also concurs. Contrarily, Fewkes believes that the Caribs did not gain a foothold there, chiefly, it seems, because the typical stonework of the Lesser Antilles associated with the Caribs by Fewkes and other archeologists has not been found in Trinidad. There is a question, however, as to whether the stone artifacts referred to are actually Carib remains. In either event, there seems to be a sufficiency of historical evidence to demonstrate the reasonableness of Loven's theory. Loven, S., *Origins of the Tainan Culture, West Indies*, Göteborg, pp. 32-42 (1935); Bullbrook, J. A., "The Ierian Race," *Public Lectures . . . of the Hist. Soc. of Trinidad and Tobago, 1938-1939* pp. 40-41 (1940); Fewkes, J. W., "A Prehistoric Island Culture Area of America," *Bur. Amer. Ethnol.*, An. Rep. 34, 64-65 (1922).

<sup>7</sup> Rainey, F. G., "A New Prehistoric Culture in Puerto Rico," *Proc. Nat. Acad. Sci.*, 21, 12-16 (1935).

<sup>8</sup> Rouse, I., "New Evidence Pertaining to Puerto Rican Prehistory," *Proc. Nat. Acad. Sci.*, 23, 185 (1937). For a discussion of the evidence that these cultures are Arawak, see Rouse's section on the West Indies, soon to be published in the "Handbook of the South American Indians," *Bur. Amer. Ethnol.*, Washington, D. C.

<sup>9</sup> Rainey, F. G., "Scientific Survey of Porto Rico and the Virgin Islands," *New York Acad. Sci.*, 18, part 1 (1940).

<sup>10</sup> Harrington, M. R., "Cuba Before Columbus," *Indian Notes and Monographs*, part 1, 2, 423 (1921); Loven, S., *Origins of the Tainan Culture, West Indies*, Göteborg, p. 662 (1935).

<sup>11</sup> I am indebted to Capt. J. E. L. Carter for unpublished information on the distinctions between the upper and lower strata of the Erin Bay site.

<sup>12</sup> Rainey, F. G., "Scientific Survey of Porto Rico and the Virgin Islands," *New York Acad. Sci.*, 18, part 1, 175 (1940).

<sup>13</sup> Our understanding of this problem will be greatly advanced by stratigraphic analyses of the Erin Bay and Barrancas deposits. The Trinidad research, it is hoped, will soon be furthered by coöperation between Trinidad archeologists and the Yale Peabody Museum.

---

## GUANINE AND FACTOR Z<sub>1</sub>, GROWTH SUBSTANCES FOR PHYCOMYCES

BY WILLIAM J. ROBBINS AND FREDERICK KAVANAGH

THE NEW YORK BOTANICAL GARDEN AND DEPARTMENT OF BOTANY, COLUMBIA UNIVERSITY

Communicated November 25, 1941

Previous reports<sup>1</sup> have shown that the spore germination, early mycelial growth and gametic reproduction of *Phycomyces* at or near 25°C. in the presence of excess thiamin are markedly benefited by extracts from various natural products. By treatment with charcoal the natural extracts were separated into two fractions.<sup>2</sup> One of these, the Ca fraction, was the material adsorbed on charcoal from acid solution and eluted with ammoniacal acetone; the other, the D<sub>R</sub> fraction, was the filtrate from the charcoal

treated extract. Both fractions were active but their effect when combined was more than additive. To account for the observed results two factors were assumed. One, factor  $Z_1$ , was assumed in the Ca fraction; the other, factor  $Z_2$ , in the  $D_R$  fraction.

Factor  $Z_1$  was soluble in water, in aqueous alcohol and in aqueous acetone. It was thermostable to autoclaving. In electrodialysis it migrated to the cathode. It was widely distributed in natural products and was not identical with pyridoxine, lactoflavin, biotin, glutamine, pathothenic acid or para-amino-benzoic acid.<sup>3</sup>

The present paper is a progress report on further work with factor  $Z_1$ .

The presence of factor  $Z_1$  was detected and the quantity present was approximated by determining the dry weight of the mycelium of *Phycomyces* produced in 72 hours at from 26° to 26.5°C. in the dark in a solution of minerals, sugar, asparagine and thiamin. To this basal solution preparations to be tested for  $Z_1$  activity were added alone and in combination with a  $D_R$  fraction prepared from potato tubers.

We have confirmed the observation that factor  $Z_1$  migrates to the cathode in electrodialysis. Our observations indicate that such migration occurs from solutions acid to about pH 5.0. Factor  $Z_1$  was not destroyed by treatment with nitrous acid, but was destroyed by bromine and by hydrogen peroxide. Its activity was not eliminated by heating in saturated  $\text{Ba}(\text{OH})_2$  or in  $4\text{NH}_4\text{SO}_4$  for 24 hours at 115°C. It was not precipitated by  $\text{Ba}(\text{OH})_2$  in methanol. It was soluble in 90 to 95 per cent ethanol. After electrodialysis it was still adsorbable on charcoal and elutable by ammoniacal acetone.

The effect of factor  $Z_1$  was not obtained with the following mixture of vitamins or vitamin-like substances: glutamine, pimelic acid, vitamin  $K_1$ , lactoflavin, thiamin, ascorbic acid, *p*-amino benzoic acid, calcium pantothenate, nicotinamide and inositol. Oxyquinoline, urea, thiourea, thiazolidine-4-carboxylic acid and *d*-ribose were ineffective. A concentrate kindly supplied by W. H. Peterson containing a factor probably identical with folic acid of Williams was ineffective as a substitute for factor  $Z_1$ , at least after autoclaving. Choline, adenine, thymine, uracil, cytosine or xanthine was ineffective.

However, guanine in amounts of 0.1  $\mu\text{g}$ . or more acted with the  $D_R$  fraction as the concentrates of factor  $Z_1$  did. This is illustrated in table 1 where the results of 3 experiments with guanine alone and in combination with a  $D_R$  fraction from potato tubers are summarized. The  $D_R$  fraction contained 145.7 mg. of dry matter per ml. There is included in the table results with concentrates of factor  $Z_1$ . The concentrate of factor  $Z_1$  for experiment 1 was prepared by saturating a Ca fraction from potato tuber tissue with  $\text{Ba}(\text{OH})_2$  and adding sufficient ethanol to make about 60% alcohol. After 48 hours at 4°C. the precipitate was removed and discarded. Sodium

nitrate was added to the filtrate which was acidified with  $\text{H}_2\text{SO}_4$  to pH 3.0. After 4 hours at room temperature the solution was evaporated, saturated with  $\text{Ba}(\text{OH})_2$  and the precipitate removed. Excess barium was removed from the filtrate with  $\text{H}_2\text{SO}_4$  and the solution adjusted to pH 6.0. This  $Z_1$  concentrate contained 3.1 mg. of dry matter per ml. The concentrate used in experiments 2 and 3 also contained 3.1 mg. of dry matter per ml. It was prepared by acidifying a Ca fraction from potato tubers with phosphoric acid and treating with sodium nitrite overnight at room temperature. It was evaporated to small volume and alcohol added to precipitate the sodium phosphate. The precipitate was removed and the alcohol evaporated. The solution was saturated with  $\text{Ba}(\text{OH})_2$  and 2 volumes of methanol added. After standing overnight the precipitate was removed and the solution treated with sulfuric acid to remove excess barium.

TABLE 1

THE EFFECT OF GUANINE AND A CONCENTRATE OF FACTOR  $Z_1$  ALONE AND IN COMBINATION WITH A  $D_R$  FRACTION (CONTAINING FACTOR  $Z_2$ ) ON THE 72-HOUR GROWTH OF *Phycomyces* AT 26°C. IN A BASAL SOLUTION OF MINERALS, SUGAR, ASPARAGINE AND THIAMIN

ADDITIONS TO 25 ML. OF BASAL SOLUTION	DRY WEIGHT PER CULTURE OF <i>Phycomyces</i> MYCELIUM, MG.					
	EXP. 1	EXP. 1 PLUS 0.1 ML. $D_R$ FRACTION	EXP. 2	EXP. 2 PLUS 0.1 ML. $D_R$ FRACTION	EXP. 3	EXP. 3 PLUS 0.1 ML. $D_R$ FRACTION
100 $\mu\text{g.}$ guanine	22.0	91.0	21.9	149.5	3.2	65.4
10 $\mu\text{g.}$ guanine	22.0	83.0	10.9	143.0	2.2	54.6
1 $\mu\text{g.}$ guanine	8.3	76.5	2.4	89.0	0.7	37.7
Nothing	4.2	67.7	1.2	52.7	Trace	20.1
1 ml. concentrate of factor $Z_1$	5.8	17.2	27.8	101.5	3.3	21.3
0.1 ml. concentrate of factor $Z_1$	9.2	79.5	6.1	111.0	1.1	44.3
0.01 ml. concentrate of factor $Z_1$	4.6	75.0	2.2	74.5	Trace	19.7

In experiment 1 the basal solution gave a yield of 8.2 mg. The addition of 1  $\mu\text{g.}$  of guanine increased the weight to 16.6 mg. and 10 or 100  $\mu\text{g.}$  were more effective. The addition of 0.1 ml. of the  $D_R$  fraction increased the weight to 135.5 mg. and combinations of guanine and the  $D_R$  fraction gave yields up to 182.0 mg. The  $Z_1$  concentrate in experiment 1 was favorable although toxicity began to appear where 1 ml. was added per flask. Although in experiment 1 the yields in the solutions containing the  $D_R$  fraction were increased by the addition of guanine and by the two lower amounts of the  $Z_1$  concentrate, in no instance were they significantly greater than the sum of the yields obtained with the  $D_R$  fraction alone and the guanine or factor  $Z_1$  concentrate alone. In experiments 2 and 3 where the yields in the check solution were considerably less than in experiment 1, probably because of the condition of the spores used as inoculum, a more



marked action of guanine and of the factor  $Z_1$  concentrate in the presence of the  $D_R$  fraction was evident.

Yeast nucleic acid was slightly effective and acid hydrolyzed nucleic acid gave effects approximately equal to the estimated guanine content.

Others have investigated pyrimidine and purine bases as growth substances for microorganisms. Richardson<sup>4</sup> found uracil to be a growth substance for *Staphylococcus aureus* under anaerobic conditions. Oxford, Lampen and Peterson<sup>5</sup> reported uracil, cytosine, guanine, adenine, xanthine and hypoxanthine ineffective with *Cl. acetobutylicum* as a substitute for the *BY* factor. Snell and Peterson<sup>6</sup> found choline, adenine, guanine, xanthine, hypoxanthine, uracil and cytosine ineffective as substitutes for an unidentified purine-like growth substance for *Lactobacillus casei*. Stockstad<sup>7</sup> reported for *Lactobacillus casei* a growth substance which appeared to be a dinucleotide containing guanine but no adenine. Guanine and thymine partially replaced it. Adenine, hypoxanthine and xanthine were as active as guanine. Uracil or cytosine did not replace thymine. Mueller and Miller<sup>8</sup> found adenine and Stockstad's dinucleotide essential for the growth of *Cl. tetani*. However, thymine and guanine were generally ineffective. Snell and Mitchell<sup>9</sup> state that each of the purine and pyrimidine bases of nucleic acid may, under certain conditions, become limiting factors for the growth of certain lactic acid bacteria. Adenine was found to stimulate the growth of *Lactobacillus arabinosus* and *L. pentosus* and to be essential for the growth of *Streptobacillus lactis*. Uracil stimulated the growth of *L. arabinosus* and of *L. mesenteroides*. Guanine was essential for the growth of *L. mesenteroides* and thymine for *S. lactis*. They state that in general the purine and pyrimidine bases were replaceable by the corresponding oxy-derivatives. Hutchings, Bohonos, Hegsted, Elvehjem and Peterson<sup>10</sup> report evidence that a norit eluate factor believed to be the nucleotide of Stockstad is of importance in the nutrition of the chick. Guanine appears to have a higher degree of specificity for *Phycomyces* than for other organisms to which it has been reported to be of some significance.

Although guanine appears to be a growth factor for *Phycomyces* under the conditions of our experiments it is not factor  $Z_1$ . This follows because treatment with nitrous acid destroyed the activity of our sample of guanine while it did not affect that of factor  $Z_1$  concentrates.

The relation of guanine to factor  $Z_1$  is problematical. There are a number of possibilities among which the following may be mentioned. Factor  $Z_1$  may be a derivative of guanine in which case both compounds may be active or one only may be active, the cell making the necessary transformation. If one only is active we have no evidence at present as to whether the active material is guanine or factor  $Z_1$ . In any event it seems that if factor  $Z_1$  is a derivative of guanine there should be a substitution for at least one hydrogen in the amino group in the second position on the purine ring since



guanine was inactivated by treatment with nitrous acid while factor  $Z_1$  was not. On the other hand guanine and factor  $Z_1$  may be two distinct and unrelated compounds each of which is effective on the same cellular system. The relatively high degree of specificity of the known vitamins would militate against this assumption. Furthermore, guanine and factor  $Z_1$  seem to have similar stabilities toward high temperature, acids, alkalies and oxidizing agents. They may, however, be unrelated substances which affect two distinct systems in the *Phycomyces* cell. Answers to these suppositions will probably have to wait the identification of factor  $Z_1$  and its use in pure chemical form.

<sup>1</sup> Robbins, W. J., *Am. Jour. Bot.*, 26, 772-778 (1939). Robbins, W. J., *Bot. Gaz.*, 101, 428-429 (1939). Robbins, W. J., *Am. Jour. Bot.*, 27, 559-564 (1940).

<sup>2</sup> Robbins, W. J., and Hamner, K. C., *Bot. Gaz.*, 101, 912-927 (1940).

<sup>3</sup> Robbins, W. J., *Ibid.*, 102, 520-535 (1940).

<sup>4</sup> Richardson, G. M., *Biochem. Jour.*, 30, 2184 (1936).

<sup>5</sup> Oxford, A. E., Lampen, J. O., and Peterson, W. H., *Ibid.*, 34, 1588-1597 (1940).

<sup>6</sup> Snell, E. E., and Peterson, W. H., *Jour. Bact.*, 39, 273-285 (1940).

<sup>7</sup> Stockstad, E. L. R., *Jour. Biol. Chem.*, 139, 475-476 (1941).

<sup>8</sup> Mueller, J. H., and Miller, P. A., *Ibid.*, 141, 933-934 (1941).

<sup>9</sup> Snell, E. E., and Mitchell, H. K., *Proc. Nat. Acad. Sci.*, 27, 1-6 (1941).

<sup>10</sup> Hutchings, B. L., Bohonos, N., Hegsted, D. M., Elvehjem, C. A., and Peterson, W. H., *Jour. Biol. Chem.*, 140, 681-682 (1941).

## RADIOACTIVE CARBON AS AN INDICATOR OF CARBON DIOXIDE UTILIZATION. VIII. THE RÔLE OF CARBON DIOXIDE IN CELLULAR METABOLISM

BY C. B. VAN NIEL, S. RUBEN, S. F. CARSON, M. D. KAMEN AND J. W. FOSTER\*

THE HOPKINS MARINE STATION, PACIFIC GROVE. DEPARTMENT OF CHEMISTRY AND  
THE RADIATION LABORATORY, UNIVERSITY OF CALIFORNIA, BERKELEY,  
CALIFORNIA

Communicated November 19, 1941

*I. Introduction.*—Previous studies on the metabolism of non-chlorophyllous living systems in the presence of labeled  $\text{CO}_2$  ( $\text{C}^{14}\text{O}_2$  or  $\text{C}^{13}\text{O}_2$ ) have shown that the organisms convert the labeled carbon into organic compounds which may occur as excreted end-products of metabolism, as cellular constituents or both. This phenomenon has been observed with *all* the systems thus far examined<sup>1-13</sup> (*Bacterium coli*, *Propionibacterium pentosaceum*, *Methanobacterium Omelianskii*, *Methanosarcina methanica*, *Clostridium acidu-urici*, *Streptococcus faecalis*, *Acetobacter* species, *Rhizopus nigricans*, *Rhizopus* species, *Aspergillus niger*, bakers yeast, Lebedev juice, barley roots, plants in the absence of light, rat liver and pigeon liver tissue).

In a few instances the utilization of  $\text{CO}_2$  had been clearly demonstrated by methods not involving the use of labeled carbon. (Thus, the utilization of  $\text{CO}_2$  by *Methanobacterium Omelianskii*, *Propionibacterium pentosaceum* and *Bacterium coli* had been established by the work of Barker,<sup>13</sup> of Wood and Werkman<sup>14</sup> and of Woods,<sup>15</sup> respectively. Also the reduction of  $\text{CO}_2$  to  $\text{CH}_4$  by *Methanosarcina methanica* had been predicted on the basis of theoretical considerations.<sup>18</sup> But in the majority of cases where the metabolic activities of the organism or tissue lead to the net production of  $\text{CO}_2$ , it is not possible to detect a small  $\text{CO}_2$  uptake by ordinary chemical methods. Thus it remained for the tracer method to prove the well-nigh universal occurrence of  $\text{CO}_2$  utilization.

For a number of years it has also been known that in the complete absence of  $\text{CO}_2$ , growth and metabolism of diverse living systems are seriously impaired. Under experimental conditions which would insure (1) the initial presence of not more than traces of  $\text{CO}_2$ , and (2) the prompt removal of additional amounts produced during metabolism, the germination of spores and growth of various microorganisms appeared to be greatly retarded, if not altogether prevented.<sup>16-20</sup> Even such respiratory activities, as methylene blue reduction by "resting cells," were shown to be similarly dependent upon the presence of  $\text{CO}_2$ .<sup>21</sup> But until now no attempt had been made to interpret these curious results.

The studies with labeled  $\text{CO}_2$  have yielded data whose interpretation in terms of general biochemical mechanisms could be attempted. Hence it seemed desirable to investigate in how far such reactions could be of aid in elucidating the effects of  $\text{CO}_2$  on growth and metabolism. Admittedly the ideas presented in this paper bear a somewhat speculative character. Nevertheless, they are capable of experimental verification, and may suggest new methods of approach to the many complicated and, at first sight apparently unrelated problems presented by living systems. In this connection may we quote G. N. Lewis<sup>22</sup> ". . . while the sort of vague surmise which is not based upon experimental evidence nor capable of experimental test has no place in our scientific method, rational speculation must always be regarded as the advance guard of experimental science."

**II. The Possible Mechanisms of  $\text{CO}_2$  Uptake by Living Systems.**—One of the most important results of biochemical investigations of recent years has been the demonstration that enzyme-catalyzed processes are, in general, reversible. This implies that a reversal of those reactions in which  $\text{CO}_2$  is liberated may be considered as possible mechanisms by which  $\text{CO}_2$  is utilized and transformed into organic compounds. A striking example of such a process is furnished by the studies of Woods<sup>15</sup> on the formation of formic acid from  $\text{CO}_2$  and  $\text{H}_2$  by *B. coli*, in accordance with the equation:



More frequently, however,  $\text{CO}_2$  production by living organisms is ascribed to an enzymatic decomposition of keto-acids:



If this type of reaction were reversible, the utilization of  $\text{CO}_2$  would then result in the formation of keto-acids<sup>44</sup> whose fundamental rôle and manifold vicissitudes in cellular metabolism have been firmly established.

Experiments reported in a previous communication have shown that the decarboxylation of  $\text{CH}_3\text{COCOOH}$  by yeast carboxylase is indeed reversible.<sup>7</sup> Yet, the reverse reaction appeared to be much too slow to account for more than a very small fraction of the  $\text{CO}_2$  uptake observed with living systems.

Investigations on  $\text{CO}_2$  assimilation by propionic acid bacteria have suggested a similar, though not identical, method by which this compound is converted into organic substances.<sup>8, 6, 9, 23</sup> The available evidence strongly supports the following mechanism:



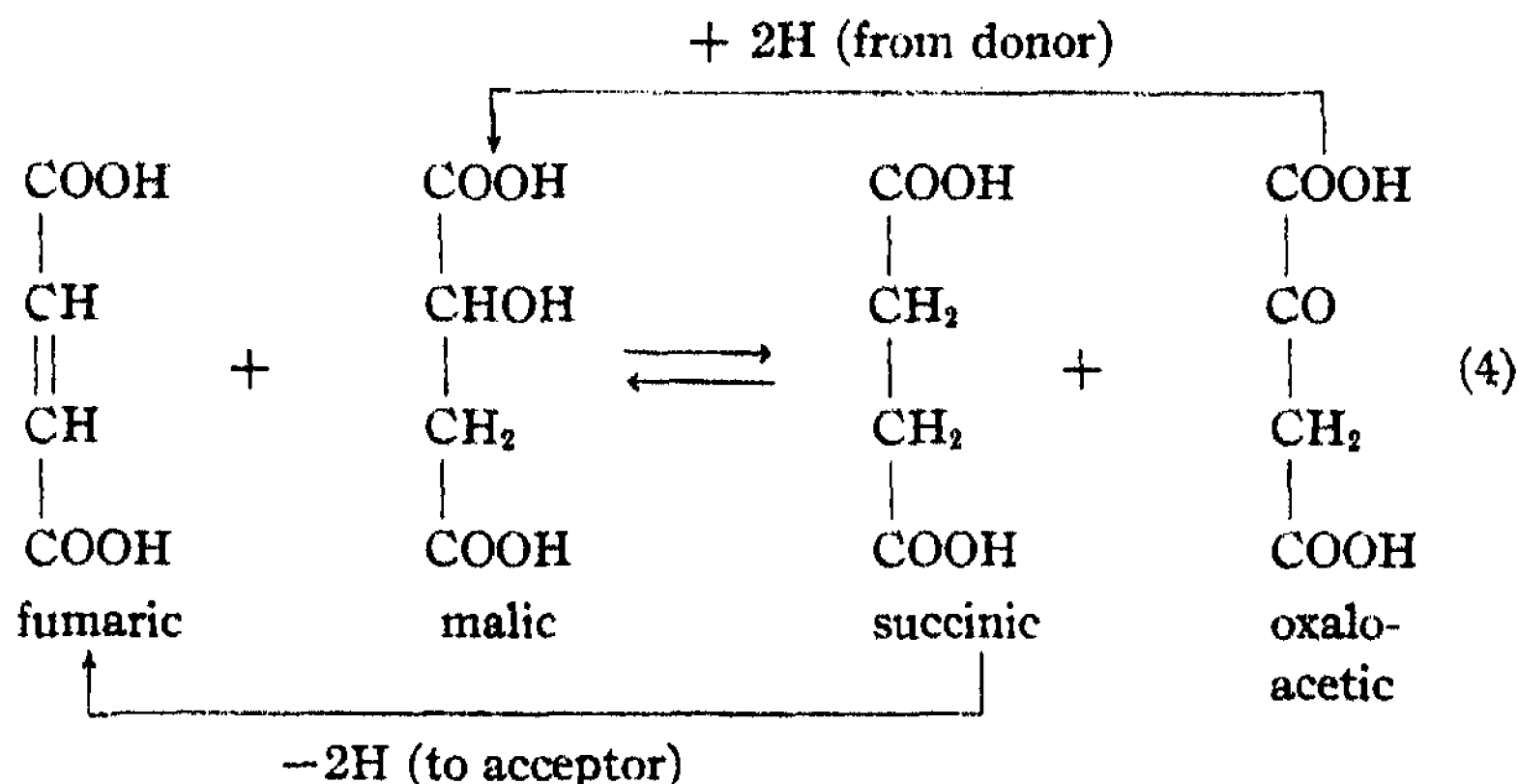
Normally, the oxaloacetic acid is reduced to succinic acid.

Furthermore, the occurrence of this reaction is apparently not limited to the propionic acid bacteria, on the contrary, it seems to be quite widespread. Various types of living systems have been shown to produce succinic and fumaric acids, at least in part, from  $\text{CO}_2$ . We refer to the work of Krebs and Eggleston with liver tissue,<sup>25</sup> of Smyth with *Staph. aureus*,<sup>26</sup> of Kleinzeller with yeast,<sup>27</sup> of Foster, *et al.*, with molds,<sup>8</sup> of van Niel, Ruben and Thomas with protozoa,<sup>28</sup> of Elsdon<sup>29</sup> and of Werkman, *et al.*,<sup>9</sup> with *B. coli*.

It is particularly the general occurrence of a  $\text{CO}_2$  fixation mechanism by which oxaloacetic acid and the closely related fumaric and succinic acids are formed which offers an opportunity for a simple interpretation of the above mentioned effects of  $\text{CO}_2$  on growth and respiration. For the studies of Szent-Györgyi and co-workers,<sup>30</sup> as well as those of numerous investigators since, have revealed that this group of  $\text{C}_4$ -acids plays a fundamental rôle in cellular metabolism.

*III. The Rôle of the  $\text{C}_4$ -Dicarboxylic Acids in Cellular Metabolism, and an Attempted Interpretation of the Necessity of  $\text{CO}_2$  for Growth and Respiration.*—According to our present knowledge in this field, it appears that the biological degradation of a substrate leads to a series of reactions in which electrons or hydrogen atoms are transferred to the ultimate acceptors by way of a rather extensive series of oxido-reduction reactions. In this chain of events, the reversible hydrogenation of oxaloacetate to malate, and of the latter (via fumarate) to succinate, plays the part of a catalytic hydrogen transporting system in various mammalian and avian tissues, as also in

microorganisms. Hence it has been proposed that the C<sub>4</sub>-dicarboxylic acids comprise a "catalytic cycle," which can be represented as follows:



This hypothesis implies that the metabolic capacity of a system would depend in part upon the presence of sufficient amounts of the participating C<sub>4</sub>-dicarboxylic acids. In growing cells the necessary increase in metabolic capacity must be brought about by a synthesis of one or more of the components of the catalytic cycle, and from the foregoing discussion it appears that such a synthesis can occur from CO<sub>2</sub> and pyruvic acid. If this were the only, or most important, way in which the C<sub>4</sub>-dicarboxylic acids *originate*, it is obvious that CO<sub>2</sub> should be an indispensable component of the medium in which growth takes place.

However, these considerations apply rigorously only to conditions under which cell multiplication occurs. Yet, it has been shown by Hes<sup>21</sup> that even in suspensions of "resting cells," CO<sub>2</sub> is necessary for the normal functioning of the respiratory mechanism which results in methylene blue reduction. The explanation of this apparent anomaly becomes clear if one bears in mind the conditions under which Hes' experiments were conducted, in particular the presence of a strong CO<sub>2</sub>-absorbing agent. The clue to an understanding of this phenomenon is furnished by the fact that oxaloacetate, apart from playing an important rôle in the Szent-Györgyi catalytic system, can as well undergo a number of other transformations. The most obvious of these is its decarboxylation. And, in the presence of a CO<sub>2</sub>-absorbing reagent, this decomposition will proceed to such an extent that the supply of this biocatalyst will gradually disappear. Consequently, it can be asserted that *the decarboxylation of oxaloacetate constitutes a "leak" through which certain essential cell constituents are drained off*. The existence of a simple mechanism by which the components of this catalytic system can be replenished by the interaction of CO<sub>2</sub> and the ubiquitous intermediate product pyruvic acid, thus becomes of great importance as a means of "plugging" this leak.

Nor is the decarboxylation of oxaloacetate the only mode of its elimination from the cycle. From the studies of Knoop,<sup>31</sup> and particularly of Virtanen, *et al.*,<sup>32</sup> one may conclude that the synthesis of dicarboxylic amino acids takes place through a reaction between oxaloacetate or fumarate and ammonia or hydroxylamine. A further possibility for loss of the components of the catalytic cycle is provided by their ready diffusibility. In order to counteract losses by any of these processes, the C<sub>4</sub>-dicarboxylic acid supply must therefore be constantly replenished.

Evans and Slotin<sup>10</sup> have recently shown that  $\alpha$ -keto glutaric acid, produced by pigeon breast muscle in the presence of C<sup>\*</sup>O<sub>2</sub>, contains radioactive carbon. In view of the work of Krebs and co-workers<sup>33</sup> it is possible to derive the ketoglutarate from oxaloacetate and pyruvate through the proposed citric acid mechanism. Also the recently reported formation of radioactive glycogen in the presence of C<sup>\*</sup>O<sub>2</sub> (Hastings, *et al.*<sup>34</sup>) falls in line with the previous considerations. As these authors point out, the most likely interpretation of their experimental results is that the C<sup>\*</sup>O<sub>2</sub> is used in the synthesis of oxaloacetate. From this latter substance, radioactive pyruvic acid can be formed, which in turn leads to the formation of radioactive carbohydrate.

Whereas these contentions point to the fundamental rôle of a reaction by which oxaloacetic acid is synthesized from CO<sub>2</sub> and pyruvic acid, it does not necessarily follow that this reaction constitutes the only important function of CO<sub>2</sub> in metabolism. It has been emphasized on account of its obvious implications for general metabolism, and also because at the present time it represents one of the possible mechanisms for CO<sub>2</sub> utilization best supported by experimental evidence. Nevertheless, it must be realized that such a formation of oxaloacetate is no more than a special case of the introduction of CO<sub>2</sub> into a molecule by means of reversible decarboxylation.

*IV. General Outlook on the Rôle of CO<sub>2</sub> in Biological Syntheses.*—That the formation of oxaloacetate is not the only way by which CO<sub>2</sub> enters into cellular metabolism is clear from a consideration of the following well-established facts:

1. The reduction of CO<sub>2</sub> to formic acid by *B. coli* (Woods<sup>15</sup>).
2. The production of CH<sub>4</sub> from CO<sub>2</sub> in the methane fermentation (Barker, *et al.*<sup>4, 18</sup>).
3. The participation of CO<sub>2</sub> in the formation of acetic acid by *Clostridium aceticum* (Wieringa<sup>35</sup>) and by *Cl. acidi-urici* (Barker, *et al.*<sup>5</sup>).
4. The utilization of CO<sub>2</sub> in the formation of urea (Krebs and Henseleit,<sup>36, 37</sup> Evans and Slotin,<sup>11</sup> Rittenberg and Waelsch.<sup>12</sup>
5. The complete synthesis of cell constituents from CO<sub>2</sub> as the only carbon source by all autotrophic organisms.

Closely related to the autotrophic organisms, from the point of view of

synthesis of cellular materials from  $\text{CO}_2$ , are those microbes which can thrive in the presence of only a single one-carbon compound ( $\text{CO}$ ,  $\text{HCOOH}$ ,  $\text{CH}_3\text{OH}$ ,  $\text{CH}_4$ ). In these cases the carbon for all the widely divergent chemical constituents comprising the cell must of necessity be ultimately derived from "one-carbon building stones." If one further remembers that there are numerous cases of facultative autotrophism in which the organism can manufacture its cell constituents either completely from  $\text{CO}_2$ , or from some simple organic substance (to this class belong all the "hydrogen bacteria," and some "sulfur bacteria") it becomes tempting to suggest the possibility that even in the presence of an organic substrate, syntheses may occur in which  $\text{CO}_2$  is one of the reactants. And in view of the fundamental similarity of the most diverse metabolic reactions,<sup>38, 39</sup> it then follows that such a possibility should also be seriously considered for the typically heterotrophic organisms.

The complex manner in which urea appears to be formed from  $\text{NH}_3$  and  $\text{CO}_2$ , indicates that the  $\text{CO}_2$  is initially built into a larger organic molecule (ornithine-citrulline-arginine cycle<sup>11, 12, 36, 37</sup>). In essence, this is somewhat similar to the above discussed mechanism for the formation of oxaloacetic acid, although not giving rise to a new carbon-carbon link. Furthermore, it is likely that also in the case of  $\text{CO}_2$  reduction to formic acid, the  $\text{CO}_2$  is first combined with an organic molecule, conceivably an enzyme, and that it is the reduction of this compound which results in the splitting off of  $\text{HCOOH}$ . The same reasoning can be applied to the formation of  $\text{CH}_4$ .

In a sense, therefore, these processes may be viewed as "syntheses of short duration." The formation of acetic acid from  $\text{CO}_2$  and  $\text{H}_2$  by *Cl. acetium* furnishes an example in which the carbon atoms from two  $\text{CO}_2$  molecules become permanently combined. Also the production of acetic acid by *Cl. acidurici* must involve such a synthesis since Barker, *et al.*, have shown that the decomposition of uric acid in the presence of  $\text{C}^*\text{O}_2$  gives rise to acetic acid in which both carbon atoms are labeled.<sup>5</sup>

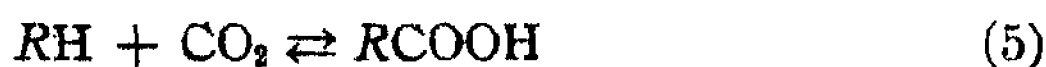
These last mentioned syntheses thus form a logical bridge to those processes in which large organic molecules are built up with the aid of  $\text{CO}_2$ . The step-wise elaboration of carbon compounds from such small elementary units would afford the most flexible mechanism for the synthesis of the endless variety of cellular constituents. Since many microorganisms are capable of effecting these syntheses starting with any one of a large number of simple carbon compounds,<sup>†</sup> it would seem entirely possible that at least for some of the syntheses  $\text{CO}_2$  is used as a building stone. Otherwise it would be necessary to postulate the existence of a large variety of synthetic mechanisms for the elaboration of the same compounds.

So far this discussion has dealt with reactions in which  $\text{CO}_2$  plays an important part, but which are independent of light. It has been deduced



from the available evidence that also in photosynthesis the actual  $\text{CO}_2$  uptake and reduction can occur in the dark (see especially the striking demonstration by Gaffron that green algae can reduce  $\text{CO}_2$  in darkness in the presence of  $\text{H}_2$ <sup>41</sup>) and that the light only serves to cause a photodecomposition of  $\text{H}_2\text{O}$ ,<sup>42</sup> thus providing for a supply of reducing substances. This tends to link the process of photosynthesis directly to all other cases of biological utilization of  $\text{CO}_2$ .

We have, in the previous section, pointed out that the formation of oxaloacetic acid from  $\text{CO}_2$  and pyruvic acid is but a special instance of a reversed decarboxylation reaction. In its most general formulation it can be represented by the equation:



The very generality of this equation renders it perhaps the most adaptable mechanism for  $\text{CO}_2$  utilization by living systems. It may well be that future work will demonstrate that the different cases of  $\text{CO}_2$  reduction are but variants of this formulation. In this connection the experiments of Ruben, *et al.*, may be mentioned.<sup>1, 43</sup> These investigators allowed green algae to utilize  $\text{C}^*\text{O}_2$  in the dark and found that the  $\text{C}^*$  assimilated became lodged in a large molecule (molecular weight  $\sim 1000$ ), and furthermore showed that a considerable portion of the  $\text{C}^*$  was present in carboxyl groups. Equation (5) was proposed to account for their results.

The general inferences that can be drawn from the preceding considerations lead, we believe, to a unified concept of the rôle played by carbon dioxide in cellular metabolism.

It is a pleasure to express our thanks to Professor H. A. Barker for many discussions and helpful suggestions.

\* National Research Council Fellow, 1939-1940. Present address: Research and Development, Merck and Co., Rahway, N. J.

† E.g., den Dooren de Jong has demonstrated that *Pseudomonas putida* can be grown with any one of some 77 different organic compounds as the sole carbon source, including such diverse molecular species as: saturated and unsaturated fatty acids, hydroxy acids, dibasic and tribasic acids, alcohols, carbohydrates, amines, amino acids, amides, aromatic compounds, etc.<sup>40</sup>

<sup>1</sup> Ruben, Kamen, Hassid and DeVault, *Science*, **90**, 570 (1939).

<sup>2</sup> Ruben and Kamen, *Proc. Nat. Acad. Sci.*, **26**, 418 (1940).

<sup>3</sup> Carson and Ruben, *Ibid.*, **26**, 422 (1940).

<sup>4</sup> Barker, Ruben and Kamen, *Ibid.*, **26**, 426 (1940).

<sup>5</sup> Barker, Ruben and Beck, *Ibid.*, **26**, 477 (1940).

<sup>6</sup> Carson, Foster, Ruben and Barker, *Ibid.*, **27**, 229 (1941).

<sup>7</sup> Carson, Ruben, Kamen and Foster, *Ibid.*, **27**, 475 (1941).

<sup>8</sup> Foster, Carson, Ruben and Kamen, *Ibid.*, **27**, 590 (1941).

<sup>9</sup> Wood, Werkman, Hemingway and Nier, *Jour. Biol. Chem.*, **135**, 789 (1940). Wood, Werkman, Hemingway and Nier, *Ibid.*, **139**, 365 (1941).

<sup>10</sup> Evans and Slotin, *Ibid.*, **136**, 301 (1940).

- <sup>11</sup> Evans and Slotin, *Ibid.*, 136, 805 (1940).  
<sup>12</sup> Rittenberg and Waelsch, *Ibid.*, 136, 799 (1940).  
<sup>13</sup> Barker, *Arch. Mikrobiol.*, 7, 404 (1936).  
<sup>14</sup> Wood and Werkman, *Biochem. Jour.*, 30, 48 (1936).  
<sup>15</sup> Woods, *Ibid.*, 30, 515 (1936).  
<sup>16</sup> Gladstone, Fildes and Richardson, *Brit. Jour. Exp. Path.*, 16, 335 (1935).  
<sup>17</sup> Rockwell and Highberger, *Jour. Infect. Dis.*, 40, 438 (1927).  
<sup>18</sup> Valley and Rettger, *Jour. Bact.*, 14, 101 (1927).  
<sup>19</sup> Rippel and Bortels, *Biochem. Z.*, 184, 237 (1927). Rippel and Heilmann, *Arch. f. Mikrobiol.*, 1, 119 (1930).  
<sup>20</sup> Longworth and MacInnes, *Jour. Bact.*, 31, 287 (1936). Longworth and MacInnes, *Ibid.*, 32, 567 (1936).  
<sup>21</sup> Hes, *Nature*, 141, 647 (1938). Hes, *Ann. d. Ferm.*, 4, 547 (1938).  
<sup>22</sup> Lewis, *Pub. Astron. Soc. Pacific*, 34, 309 (1922).  
<sup>23</sup> Krebs and Eggleston, *Biochem. Jour.*, 35, 676 (1941).  
<sup>24</sup> The beta-decarboxylation of oxaloacetic acid, heretofore purely hypothetical, has now been convincingly demonstrated by Krampitz and Werkman, *Ibid.*, 35, 595 (1941).  
<sup>25</sup> Krebs and Eggleston, *Ibid.*, 34, 1383 (1940).  
<sup>26</sup> Smyth, *Ibid.*, 34, 1598 (1940).  
<sup>27</sup> Kleinzeller, *Ibid.*, 35, 495 (1941).  
<sup>28</sup> van Niel, Ruben and Thomas (unpublished).  
<sup>29</sup> Elsdon, *Biochem. Jour.*, 32, 187 (1938).  
<sup>30</sup> Szent-Györgyi, *Arch. exp. Zellforsch.*, 15, 29 (1934). Szent-Györgyi, *Nature*, 135, 305 (1935). Banga, Gerendás, Laki, Papp, Porges, Straub and Szent-Györgyi, *Z. physiol. Chem.*, 254, 147 (1937). Szent-Györgyi, *Acta. Med. Szeged*, 9, 1 (1937). Szent-Györgyi, *On Oxidation, Fermentation, Vitamins, Health and Disease*, Williams and Wilkins Co., Baltimore, 1939.  
<sup>31</sup> Knoop, *Oxydationen im Tierkörper*, F. Enke, Stuttgart, 1931.  
<sup>32</sup> Virtanen, *Cattle Fodder and Human Nutrition*, Univ. Press, Cambridge, 1938. Virtanen and Tarnanen, *Biochem. Z.*, 250, 193 (1932). Virtanen and Laine, *Suomen Kemistilehti*, B-9, 12 (1936).  
<sup>33</sup> Krebs, *Nature*, 138, 288 (1936). Krebs, *Biochem. Jour.*, 34, 775 (1940). Krebs and Eggleston, *Ibid.*, 34, 1383 (1940). Krebs and Johnson, *Enzymologia*, 4, 148 (1937). Krebs, Eggleston, Kleinzeller and Smyth, *Biochem. J.*, 34, 1234 (1940). Evans, *Ibid.*, 34, 829 (1940).  
<sup>34</sup> Cramer and Kistiakowsky, *Jour. Biol. Chem.*, 137, 549 (1941). Conant, Kramer, Hastings, Klemperer, Solomon and Vennesland, *Ibid.*, 137, 557 (1941). Solomon, Vennesland, Klemperer, Buchanan and Hastings, *Ibid.*, 140, 171 (1941).  
<sup>35</sup> Wieringa, *Ant. van Leeuwenhoek*, 3, 88 (1936). Wieringa, *Ibid.*, 3, 263 (1936). Wieringa, *Ibid.*, 6, 251 (1941).  
<sup>36</sup> Krebs and Henseleit, *Z. physiol. Chem.*, 210, 33 (1932).  
<sup>37</sup> Krebs, *Ibid.*, 217, 191 (1933). Krebs, *Ergeb. d. Enz.*, 3, 247 (1934).  
<sup>38</sup> Kluyver and Donker, *Chem. Zelle u. Gewebe*, 13, 134 (1926).  
<sup>39</sup> Kluyver, *The Chemical Activities of Micro-Organisms*, Univ. of London Press, 1931.  
<sup>40</sup> den Dooren de Jong, "Bijdrage tot de kennis van het mineralisatieproces." Dissertation, Rotterdam (1926).  
<sup>41</sup> Gaffron, *Science*, 91, 529 (1940).  
<sup>42</sup> van Niel, *Advances in Enzymology*, 1, 263 (1941).  
<sup>43</sup> Ruben, Hassid and Kamen, *Jour. Am. Chem. Soc.*, 61, 661 (1939). Ruben, Kamen and Hassid, *Ibid.*, 62, 3443 (1940). Ruben, Kamen and Perry, *Ibid.*, 62, 3450 (1940). Ruben and Kamen, *Ibid.*, 62, 3451 (1940).  
<sup>44</sup> Thimann, *Science*, 88, 506 (1938).



# REPRESENTATION OF MEASURABLE FLOWS IN EUCLIDEAN 3-SPACE

BY SHIZUO KAKUTANI

INSTITUTE FOR ADVANCED STUDY

Communicated November 19, 1941

1. The purpose of this note is to show that every measurable flow defined on a measure space satisfying certain separability conditions is isomorphic to a continuous flow on a subset of a Euclidean 3-space with ordinary Lebesgue measure.

2. We begin with definitions.<sup>1</sup> A *measure space*  $\Omega(\mathfrak{B}, m)$  is a system of a space  $\Omega$ , a Borel field  $\mathfrak{B}$  of subsets  $M$  of  $\Omega$ , and a countably additive completed measure  $m(M)$  defined on  $\mathfrak{B}$  such that  $0 < m(\Omega) < \infty$ . A set belonging to  $\mathfrak{B}$  is called *measurable*. A measure space  $\Omega(\mathfrak{B}, m)$  is *properly separable* if there exists a countable collection  $\mathfrak{A}$  of measurable sets, called a *basis*, such that, for any  $M \in \mathfrak{B}$  and for any  $\epsilon > 0$ , there exists a subsequence

$\{M_n\}$  ( $n = 1, 2, \dots$ ) from  $\mathfrak{A}$  such that  $M \subset \sum_{n=1}^{\infty} M_n$  and  $\sum_{n=1}^{\infty} m(M_n) < m(M) + \epsilon$ .

A countable collection  $\mathfrak{A}$  of subsets of  $\Omega$  is a *separating sequence* of  $\Omega$  if, for any two distinct points of  $\Omega$ , there exists at least one set from  $\mathfrak{A}$  which contains one but not both of them.

If  $\Omega$  is a Lebesgue measurable subset of a Euclidean  $n$ -space  $R^n$  with a finite Lebesgue measure, and if  $\mathfrak{L}(\Omega)$  is the collection of all Lebesgue measurable subsets of  $\Omega$ , then  $\Omega(\mathfrak{L}(\Omega), \mu)$  ( $\mu$  means  $n$ -dimensional Lebesgue measure in  $R^n$ ) is called a *Lebesgue measure space*. More generally, if  $\Omega$  is an arbitrary subset of a Euclidean  $n$ -space  $R^n$  with a finite Lebesgue outer measure, and if  $\mathfrak{L}^*(\Omega)$  is the collection of all subsets  $M$  of  $\Omega$  of the form:  $M = \Omega \cdot \Lambda$ , where  $\Lambda$  is a Lebesgue measurable set of  $R^n$ , then  $\Omega(\mathfrak{L}^*(\Omega), \mu^*)$  ( $\mu^*$  means  $n$ -dimensional Lebesgue outer measure in  $R^n$ ) is called a *Lebesgue\* measure space*.<sup>2</sup> Clearly, Lebesgue and Lebesgue\* measure spaces are properly separable and have a separating sequence of measurable sets.

A *measure preserving transformation* (of a measure space onto another or onto the same measure space), a *flow* (= one parameter group of measure preserving transformations of a measure space onto itself), and the notions of *invariant sets* and *ergodicity* are defined as usual. A flow  $\{T_t\}$  ( $-\infty < t < +\infty$ ) defined on a measure space  $\Omega(\mathfrak{B}, m)$  is *measurable* if, for any  $M \in \mathfrak{B}$ , the set  $\bar{M} = [(\omega, t) : T_t(\omega) \in M]$  is measurable in the product space of  $\Omega$  with the real  $t$ -axis, where measure is defined multiplicatively in terms of the given measure  $m(M)$  on  $\Omega$  and the Lebesgue measure on  $t$ -axis. Two flows  $\{S_t\}$  and  $\{T_t\}$  defined on two measure spaces  $\Omega(\mathfrak{B}, m)$  and  $\Omega'(\mathfrak{B}', m')$ , respectively, are *isomorphic* to each other if it is possible to find two in-

variant sets of measure zero  $N \in \mathfrak{B}$ ,  $N' \in \mathfrak{B}'$  and a measure preserving transformation of  $\Omega - N$  onto  $\Omega' - N'$  which carries  $\{S_i\}$  into  $\{T_i\}$ .

3. Our main results are:

**THEOREM 1.** *Let  $\{T_i\}$  be a measurable flow defined on a properly separable measure space having a separating sequence of measurable sets. If every point of the space is of measure zero, then  $\{T_i\}$  is isomorphic to a continuous flow on a Lebesgue\* measure space in a Euclidean 3-space  $R^3$ .*

**THEOREM 2.** *Every measurable flow defined on a Lebesgue measure space in  $R^n$  ( $n \geq 1$ ) is isomorphic to a continuous flow on a Lebesgue measure space in  $R^3$ .*

It is desirable to represent  $\{T_i\}$  as a differentiable or an analytic flow in  $R^3$ . But this is not always possible. For instance, if  $\{T_i\}$  is a group of rotations of a circumference (with one-dimensional Lebesgue measure), then the corresponding representation space in  $R^3$  must be a simple Jordan curve with a positive three-dimensional Lebesgue measure. It is, however, easy to see that such a curve cannot be differentiable. The proof of Theorem 1 will be given in sections 5, 6 and 7, and Theorem 2 will be proved in section 6.

4. Let  $\Omega(\mathfrak{B}, m)$  be a measure space, and let  $S$  be a measure preserving transformation of  $\Omega(\mathfrak{B}, m)$  onto itself. Let further  $f(\omega)$  be a positive valued measurable and integrable function defined on  $\Omega$ . We assume that there exists a constant  $c > 0$  such that  $f(\omega) \geq c$  for all  $\omega \in \Omega$ . Consider the product space of  $\Omega$  with the real  $u$ -axis, with the measure  $\bar{m}(\bar{M})$  defined on it multiplicatively in terms of the measure  $m(M)$  on  $\Omega$  and the Lebesgue measure on the  $u$ -axis. Let  $\bar{\Omega}$  be the portion of this product space under the graph of  $f(\omega)$ , i.e. the set of all points of the form  $\bar{\omega} = (\omega, u)$ ,  $\omega \in \Omega$ ,  $0 \leq u < f(\omega)$ , and let  $\bar{\mathfrak{B}}$  be the collection of all  $\bar{m}$ -measurable subsets  $\bar{M}$  of  $\bar{\Omega}$ . Then  $\bar{\Omega}(\bar{\mathfrak{B}}, \bar{m})$  is a measure space and the flow  $\{T_i\}$  defined on it by<sup>3</sup>

$$\begin{aligned} T_i(\omega, u) &= (\omega, u + t), \text{ if } -u \leq t < -u + f(\omega), \\ &= (S^n(\omega), u + t - f(\omega) - \dots - f(S^{n-1}(\omega))), \\ &\quad \text{if } -u + \sum_{k=0}^{n-1} f(S^k(\omega)) \leq t < -u + \sum_{k=0}^n f(S^k(\omega)), n = 1, 2, \dots, \\ &= (S^{-n}(\omega), u + t + f(S^{-1}(\omega)) + \dots + f(S^{-n}(\omega))), \\ &\quad \text{if } -u - \sum_{k=1}^n f(S^{-k}(\omega)) \leq t < -u - \sum_{k=1}^{n-1} f(S^{-k}(\omega)), n = 1, 2, \dots, \end{aligned}$$

is called a *flow built under a function*.  $f(\omega)$  is called the *ceiling function*,  $\Omega$  and  $S$  are called the *base space* and the *base transformation*, respectively.

5. It was proved by W. Ambrose<sup>4</sup> that every ergodic measurable flow (defined on an arbitrary measure space) is isomorphic to a flow built under a function. Later it was shown<sup>5</sup> that the same thing is true for general non-ergodic measurable flows (defined on a measure space having a separating sequence of measurable sets) if we permit that the base space has an infinite

measure. More precisely speaking, given a measurable flow  $\{T_t\}$  defined on a measure space  $\Omega(\mathfrak{B}, m)$  having a separating sequence of measurable sets, it is possible to decompose  $\Omega$  into a countable number of disjoint invariant measurable sets:

$\Omega = \sum_{n=0}^{\infty} \Omega_n$  (some of the terms may be missing) in such a way that (i)  $\{T_t\}$  is the identity on  $\Omega_0$ , and (ii)  $\{T_t\}$  is isomorphic to a flow built under a function on each  $\Omega_n$  ( $n = 1, 2, \dots$ ).

Thus the proof of Theorem 1 is reduced to the discussion of the special cases (i) and (ii). Since the first case can be easily discussed,<sup>6</sup> we shall discuss only the second case. Hence, from now on, we assume that  $\{T_t\}$  is a flow built under a function, and we adopt the notation of section 4.

6. By assumption, there exists a constant  $c > 0$  such that  $f(\omega) \geq c$  for all  $\omega \in \Omega$ . We may also assume that  $f(\omega) \leq 2c$  for all  $\omega \in \Omega$ . Next we remark that the base space  $\Omega(\mathfrak{B}, m)$  is properly separable and has a separating sequence of measurable sets.<sup>7</sup> Let  $\mathfrak{F} = \{M_n\}$  ( $n = 1, 2, \dots$ ) be a basis of  $\Omega(\mathfrak{B}, m)$  which is at the same time a separating sequence. We may assume that  $\mathfrak{F}$  contains all sets  $M$  of the form:  $M = [\omega: a \leq f(\omega) < b]$ , where  $a$  and  $b$  are arbitrary rational numbers. Further we may assume that  $\mathfrak{F}$  is invariant under the base transformation  $S$  (i.e.,  $M \in \mathfrak{F}$  implies  $S(M) \in \mathfrak{F}$  and  $S^{-1}(M) \in \mathfrak{F}$ ). Finally, we may assume that  $\mathfrak{F}$  is a finite field.

Let us now define the distance of two points  $\omega$  and  $\omega'$  of  $\Omega$  by  $d(\omega, \omega') = \sum_{n=1}^{\infty} 3^{-n} |\varphi_n(\omega) - \varphi_n(\omega')|$ , where  $\varphi_n(\omega)$  is the characteristic function of  $M_n$ .

It is easy to see that, with respect to this metric,  $\Omega$  is a zero-dimensional separable metric space and that the ceiling function  $f(\omega)$  is a continuous function. Moreover,  $\omega' = S(\omega)$  is a homeomorphism of  $\Omega$  onto itself, and the following conditions are satisfied:

- (1) every open set (and hence every Borel set) is measurable,
- (2) every measurable set is contained in a Borel set of the same measure.

Let us consider the set  $\bar{\Omega}^0$  of all points  $\bar{\omega}$  of the form:  $\bar{\omega} = (\omega, u)$ ,  $\omega \in \Omega$ ,  $0 \leq u \leq f(\omega)$ , and define the distance of two points  $\bar{\omega} = (\omega, u)$  and  $\bar{\omega}' = (\omega', u')$  of  $\bar{\Omega}^0$  by  $\bar{d}(\bar{\omega}, \bar{\omega}') = d(\omega, \omega') + |u - u'|$ . Then  $\bar{\Omega}^0$  is clearly a one-dimensional separable metric space. If we identify two points  $(\omega, f(\omega))$  and  $(S(\omega), 0)$  of  $\bar{\Omega}^0$  for each  $\omega \in \Omega$ , then we have a new metric space. This may be observed as a metrization of  $\bar{\Omega}$ , and it is not difficult to see that, with respect to this metrization,  $\bar{\Omega}$  is a one-dimensional separable metric space, on which  $\{T_t\}$  is a continuous flow. It is also not difficult to see that the conditions (1) and (2) are satisfied by the measure space  $\bar{\Omega}(\bar{\mathfrak{B}}, \bar{m})$ . Moreover, by assumption,

- (3) every point is of measure zero;

but it is not necessarily true that

- (4) every open set is of positive measure.

Let  $\bar{N}$  be the union of all open sets of zero measure of  $\bar{\Omega}$ , if there are any. Then  $\bar{N}$  is an invariant null set of measure zero, and the remaining part  $\bar{\Omega} - \bar{N}$  clearly satisfies the conditions (1), (2), (3) and (4). Hence we may assume that  $\bar{\Omega}$  itself satisfies all these conditions.

7. By a well-known theorem,<sup>8</sup> we can embed  $\bar{\Omega}$  homeomorphically into a bounded part of a Euclidean 3-space  $R^3$ , and thus we obtain a representation of the given flow as a continuous flow defined on a subset of  $R^3$ . The proof of our theorem, however, is not yet finished since the measure space  $\bar{\Omega}(\bar{\mathfrak{B}}, \bar{m})$ , thus embedded into  $R^3$ , does not necessarily coincide with the Lebesgue\* measure space  $\bar{\Omega}(\mathfrak{L}^*(\bar{\Omega}), \mu^*)$ .

Let  $\bar{\Omega}^*$  be a bounded Borel set of  $R^3$  such that  $\bar{\Omega} \subset \bar{\Omega}^*$  and  $\mu^*(\bar{\Omega}) = \mu(\bar{\Omega}^*)$ , where  $\mu$  and  $\mu^*$  mean Lebesgue and Lebesgue outer measures in  $R^3$ . Let  $Q$  be a closed cube in  $R^3$  which contains  $\bar{\Omega}^*$  entirely in its interior, and let  $\bar{\mathfrak{B}}^*$  be the collection of all subsets  $M$  of  $Q$  such that  $M \cdot \bar{\Omega} \in \bar{\mathfrak{B}}$  and  $M - M \cdot \bar{\Omega}^*$  is Lebesgue measurable.  $\bar{\mathfrak{B}}^*$  is clearly a Borel field, and if we put  $\nu(M) = \bar{m}(M \cdot \bar{\Omega}) + \mu(M - M \cdot \bar{\Omega}^*)$ , then  $\nu(M)$  is a countably additive completed measure defined on  $\bar{\mathfrak{B}}^*$ . It is clear that the condition (1) is satisfied by the measure space  $Q(\bar{\mathfrak{B}}^*, \nu)$ , but the condition (2) is not necessarily satisfied.

Let  $\bar{\mathfrak{B}}^{**}$  be the collection of all subsets  $\Lambda$  of  $Q$  such that there exist two Borel sets  $\Lambda_1$  and  $\Lambda_2$  with  $\Lambda_1 \subset \Lambda \subset \Lambda_2 \subset Q$  and  $\nu(\Lambda_1) = \nu(\Lambda) = \nu(\Lambda_2)$ . Then  $\bar{\mathfrak{B}}^{**}$  is a Borel subfield of  $\bar{\mathfrak{B}}^*$ , and the measure space  $Q(\bar{\mathfrak{B}}^{**}, \nu)$  clearly satisfies the conditions (1) and (2). Since the conditions (3) and (4) are clearly satisfied by the measure space  $Q(\bar{\mathfrak{B}}^{**}, \nu)$ , and since the boundary of  $Q$  is a set of  $\nu$ -measure zero, there exists, by a theorem of J. C. Oxtoby-S. M. Ulam<sup>9</sup> and J. von Neumann, a homeomorphism  $\omega' = h(\omega)$  of  $Q$  onto another closed cube  $R$ , which carries the measure space  $Q(\bar{\mathfrak{B}}^{**}, \nu)$  into the Lebesgue measure space  $R(\mathfrak{L}(R), \mu)$ .

Since  $Q(\bar{\mathfrak{B}}^*, \nu)$  is an extension of  $Q(\bar{\mathfrak{B}}^{**}, \nu)$ , it will be carried over by this homeomorphism  $\omega' = h(\omega)$  into a measure space which is an extension of the Lebesgue measure space  $R(\mathfrak{L}(R), \mu)$ . We denote this measure space by  $R(\bar{\mathfrak{L}}(R), \bar{\mu})$ , where  $\bar{\mu}$  is an extension of the ordinary Lebesgue measure defined on a Borel field  $\bar{\mathfrak{L}}(R)$  containing all Lebesgue measurable subsets of  $R$ . Let  $\bar{\Omega}' = h(\bar{\Omega})$  be the image of  $\bar{\Omega}$  by  $h(\omega)$ , and let us consider the part of  $R(\bar{\mathfrak{L}}(R), \bar{\mu})$  restricted on  $\bar{\Omega}'$ , which we shall denote by  $\bar{\Omega}'(\bar{\mathfrak{L}}(\bar{\Omega}'), \bar{\mu})$ . Since it is clear that the measure space  $\bar{\Omega}'(\bar{\mathfrak{L}}(\bar{\Omega}'), \bar{\mu})$  is isomorphic to the given measure space  $\bar{\Omega}(\bar{\mathfrak{B}}, \bar{m})$ , the proof of our theorem will be finished if we can prove that  $\bar{\Omega}'(\bar{\mathfrak{L}}(\bar{\Omega}'), \bar{\mu})$  is the same measure space as the Lebesgue\* measure space  $\bar{\Omega}'(\mathfrak{L}^*(\bar{\Omega}'), \mu^*)$ .

For this purpose, let  $M' \in \mathfrak{L}^*(\bar{\Omega}')$ . Then, by definition, there exists a Lebesgue measurable set  $\Lambda'$  such that  $M' = \Lambda' \cdot \bar{\Omega}'$ . We may assume that  $\Lambda' \subset \bar{\Omega}^{*'} = h(\bar{\Omega}^*)$ , and  $\mu^*(M') = \mu(\Lambda')$ . Let  $\Lambda'_1$  and  $\Lambda'_2$  be two Borel sets such that  $\Lambda'_1 \subset \Lambda' \subset \Lambda'_2 \subset \bar{\Omega}^{*'}$  and  $\mu(\Lambda'_1) = \mu(\Lambda') = \mu(\Lambda'_2)$ . Let further  $M$ ,  $\Lambda_1$ ,  $\Lambda$  and  $\Lambda_2$  be the inverse images of  $M'$ ,  $\Lambda'_1$ ,  $\Lambda'$  and  $\Lambda'_2$  by  $h(\omega)$ .  $\Lambda_1$  and  $\Lambda_2$

are Borel sets, and we have  $\nu(\Lambda_1 - \Lambda_2) = 0$  since  $\mu(\Lambda'_1 - \Lambda'_2) = 0$ . Since  $\Lambda_1 \subset \Lambda \subset \Lambda_2$ , we have  $\Lambda \in \mathfrak{B}^{**}$ , and consequently  $M = \Lambda \cdot \bar{\Omega} \in \mathfrak{B}^*$ , which in turn implies that  $M' \equiv h(M) \in \bar{\mathfrak{X}}(\bar{\Omega}')$ . Thus we have proved that  $\mathfrak{X}^*(\bar{\Omega}') \subset \bar{\mathfrak{X}}(\bar{\Omega}')$ . Moreover, we have  $\mu^*(M') = \mu(\Lambda') = \nu(\Lambda) \geq \nu(M) = \bar{m}(M) = \bar{\mu}(M')$ .

Conversely, let  $M' \in \bar{\mathfrak{X}}(\bar{\Omega}')$ . Then there exists an  $M \subset \Omega$  such that  $M' = h(M)$  and  $M \in \mathfrak{B}^*$ . Since  $M \subset \bar{\Omega}$ , we must have  $M \in \mathfrak{B}$ . Hence there exist two Borel sets  $\Lambda_1$  and  $\Lambda_2$  such that  $\Lambda_1 \cdot \bar{\Omega} \subset M \subset \Lambda_2 \cdot \bar{\Omega}$  and  $\bar{m}(\Lambda_1 \cdot \bar{\Omega}) = \bar{m}(M) = \bar{m}(\Lambda_2 \cdot \bar{\Omega})$ . We may assume that  $\Lambda_1 \subset \Lambda_2 \subset \bar{\Omega}^*$ . Let  $\Lambda$  be an arbitrary set such that  $\Lambda_1 \subset \Lambda \subset \Lambda_2$  and  $M = \Lambda \cdot \bar{\Omega}$ . Let further  $\Lambda'_1$ ,  $\Lambda'$  and  $\Lambda'_2$  be the images of  $\Lambda_1$ ,  $\Lambda$  and  $\Lambda_2$  by  $h(\omega)$ . Then  $\Lambda'_1$  and  $\Lambda'_2$  are Borel sets, and, since  $\nu(\Lambda_2 - \Lambda_1) = \bar{m}((\Lambda_2 - \Lambda_1) \cdot \bar{\Omega}) = 0$ , we have  $\mu(\Lambda'_2 - \Lambda'_1) = 0$ . Since  $\Lambda' \subset \Lambda'_1 \subset \Lambda'_2$ , we have  $\Lambda' \in \mathfrak{X}(Q)$ , which together with the relation  $M' = \Lambda' \cdot \bar{\Omega}'$  will imply  $M' \in \mathfrak{X}^*(\bar{\Omega}')$ . Thus we have proved that  $\bar{\mathfrak{X}}(\bar{\Omega}') \subset \mathfrak{X}^*(\bar{\Omega}')$ . Moreover, we have  $\bar{\mu}(M') = \bar{m}(M) = \nu(\Lambda) = \mu(\Lambda') \geq \mu^*(M')$ . Combined with the inequality obtained above, this will give  $\bar{\mu}(M') = \mu^*(M')$  for all  $M' \in \bar{\mathfrak{X}}(\bar{\Omega}') = \mathfrak{X}^*(\bar{\Omega}')$ , and the proof of  $\bar{\Omega}'(\bar{\mathfrak{X}}(\bar{\Omega}'), \bar{\mu}) = \bar{\Omega}'(\mathfrak{X}^*(\bar{\Omega}'), \mu^*)$  is completed.

8. In order to prove Theorem 2, we need only prove the following<sup>10</sup>:

LEMMA. If a Lebesgue\* measure space  $\Omega^*(\mathfrak{X}^*(\Omega^*), \mu^*)$  in  $R^m$  ( $m \geq 1$ ) is isomorphic to a Lebesgue measure space  $\Omega(\mathfrak{X}(\Omega), \mu)$  in  $R^n$  ( $n \geq 1$ ), then  $\Omega^*$  is a Lebesgue measurable subset of  $R^m$ .

Proof. Let  $\omega^* = \varphi(\omega)$  be a measure preserving transformation of  $\Omega(\mathfrak{X}(\Omega), \mu)$  onto  $\Omega^*(\mathfrak{X}^*(\Omega^*), \mu^*)$  which gives this isomorphism.  $\varphi(\omega)$  is clearly Lebesgue measurable. Hence there exists a Borel measurable mapping  $\psi(\omega)$  such that  $\varphi(\omega) = \psi(\omega)$  almost everywhere on  $\Omega$ . Let  $\Omega_0$  be a Borel set such that  $\Omega_0 \subset \Omega$ ,  $\mu(\Omega - \Omega_0) = 0$  and  $\varphi(\omega) = \psi(\omega)$  everywhere on  $\Omega_0$ . Then  $\Omega^* = \varphi(\Omega) = \varphi(\Omega_0) + \varphi(\Omega - \Omega_0) = \psi(\Omega_0) + \varphi(\Omega - \Omega_0)$ . Since  $\psi(\Omega_0)$  is Lebesgue measurable as the image of a Borel set  $\Omega_0$  by a Borel measurable mapping  $\psi(\omega)$ , and since  $\varphi(\Omega - \Omega_0)$  is of Lebesgue measure zero as the image of a set of measure zero  $\Omega - \Omega_0$  by a measure preserving transformation  $\varphi(\omega)$ ,  $\Omega^*$  must be measurable.

<sup>1</sup> For details of definition, see Ambrose, W., and Kakutani, S., "Structure and Continuity of Measurable Flows," to appear in *Duke Math. Jour.* The author is much indebted to Dr. W. Ambrose for his kind conversations on the subjects discussed in the present paper.

<sup>2</sup> See Doob, J. L., "Stochastic Process Depending on a Continuous Parameter," *Trans. Amer. Math. Soc.*, 42, 107-140 (1937).

<sup>3</sup> See Ambrose, W., "Representation of Ergodic Flows," *Ann. Math.*, 42, 723-739 (1941).

<sup>4</sup> See reference 3 above.

<sup>5</sup> See reference 1 above.

<sup>6</sup> In this case, we have only to prove that every properly separable measure space, which has a separating sequence of measurable sets and on which every point is of mea-

sure zero, is isomorphic to a Lebesgue\* measure space in  $R^3$ . We can even prove that such a measure space is isomorphic to a Lebesgue\* measure space in  $R^1$ .

<sup>7</sup> See reference 1 above, Lemmas 5 and 6.

<sup>8</sup> This embedding can be carried out in a concrete way. First we map the base space  $\Omega$  onto a zero-dimensional set  $X$  in the interval  $0 \leq x \leq 1$  by  $x = \sum 3^{-n} \phi_n(\omega)$ , where  $\phi_n(\omega)$  is the characteristic function of  $M_n$ . Then  $\bar{\Omega}^0$  can be considered as a set in the  $(x, u)$ -plane. The final embedding of  $\bar{\Omega}$  into  $R^3$  can be obtained by twisting  $\bar{\Omega}^0$  in  $R^3$  and by identifying the points which correspond to  $(\omega, f(\omega))$  and  $(S(\omega), 0)$ .

<sup>9</sup> Oxtoby, J. C., and Ulam, S. M., "Measure Preserving Homeomorphism and Metrical Transitivity," *Ann. Math.*, 42, 874-920 (1941). Theorem 2<sub>1</sub>.

<sup>10</sup> Theorem 2 is also true if the given measure space  $\Omega(\mathfrak{B}, m)$  is normal in the sense of Halmos, P., and von Neumann, J., *Bull. Amer. Math. Soc.*, 47, 696-697 (1941).

## INVERSION OF A GENERALIZED LAPLACE INTEGRAL

BY R. P. BOAS, JR.

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY

Communicated December 5, 1941

The transform

$$f(z) = \left(\frac{2z}{\pi}\right)^{1/2} \int_0^\infty t^{1/2} K_\nu(z t) d\alpha(t), \quad (1)$$

in which  $K_\nu(x)$  denotes a Bessel function of imaginary argument,<sup>1</sup> has recently been discussed by C. S. Meijer<sup>2</sup> and R. E. Greenwood.<sup>3</sup> This transform is a generalization of the Laplace transform, to which it reduces when  $\nu = \pm 1/2$ . Meijer obtained an inversion formula for (1) generalizing the complex inversion formula for Laplace transforms. The object of this note is to point out that (1) can also be inverted by means of a differential operator which is a generalization of the Post-Widder inversion operator<sup>4</sup> for Laplace transforms. In addition, necessary and sufficient conditions for the representation of a given function  $f(z)$  in the form (1) can be expressed in terms of the inversion operator.<sup>5</sup>

Let us define differential operators  $W$ ,  $Q_k$ , by the relations

$$W[g(z)] = z^{2\nu-1} [z^{1-2\nu} g'(z)]',$$

$$Q_k[f(u)] = \frac{1}{(2k)!} \left(\frac{2k}{u}\right)^{2k+3/2-\nu} W^k [z^{\nu-1/2} f(z)] \Big|_{z=2k/u}.$$

We suppose that  $-1/2 < \Re(\nu) < 1/2$ , and that (1) converges for some  $z$  (and hence for any  $z'$  with larger real part).

**THEOREM 1.** *If  $\alpha(t)$  is the indefinite integral of  $\varphi(u)$ , then for almost all positive  $u$*

$$\varphi(u) = \lim_{k \rightarrow \infty} Q_k[f(u)]. \quad (2)$$

THEOREM 2. If  $\alpha(t)$  is a function of bounded variation on every finite interval, then for all positive  $t$

$$\frac{\alpha(t+) + \alpha(t-)}{2} - \alpha(0+) = \lim_{k \rightarrow \infty} \int_0^t Q_k[f(u)] du.$$

THEOREM 3. The function  $f(z)$  has the representation (1) with  $\alpha(t)$  the integral of a bounded function if and only if  $f(x)$  is of class  $C^\infty$  on  $(0, \infty)$ ,  $f(\infty) = 0$ , and

$$|Q_k[f(u)]| \leq M \quad (0 < u < \infty; k = 1, 2, \dots).$$

Other representation theorems, resembling known theorems for Laplace transforms,<sup>6</sup> can also be proved. However, it seems unlikely that  $Q_k[f(u)] \geq 0$  is either necessary or sufficient for (1) with increasing  $\alpha(t)$ , if  $\nu$  is not real.

Proofs of Theorems 1 and 3 will now be sketched.

LEMMA. If  $\varphi(u)$  is integrable on every finite interval and  $\int_0^\infty \varphi(u) du = O(e^{cu})$  for some  $c$ , then the integrals

$$I_k(u) = \left(\frac{2}{\pi}\right)^{1/2} \frac{(2k)^{2k+3/2}}{(2k)!} \int_0^\infty y^{2k+1/2} K_\nu(2ky) \varphi(uy) dy$$

exist for sufficiently large  $k$ , and  $\lim_{k \rightarrow \infty} I_k(u) = \varphi(u)$  for almost all positive  $u$ , including all points of continuity of  $\varphi(u)$ .

The hypotheses of the lemma are easily verified if (1) converges with  $(\alpha t) = \int_0^\infty \varphi(u) du$ . It is known<sup>7</sup> that for almost all positive  $u$

$$\lim_{n \rightarrow \infty} \frac{n^{n+1}}{n!} \int_0^\infty y^n e^{-ny} \{\varphi(uy) - \varphi(u)\} dy = 0. \quad (4)$$

The lemma is proved by combining (4) with estimates furnished by the asymptotic expansion<sup>8</sup> of  $K_\nu(x)$ .

From the elementary properties<sup>9</sup> of  $K_\nu(x)$  we find that

$$W_z[z^\nu K_\nu(zt)] = t^2 z^\nu K_\nu(zt). \quad (5)$$

It is easy to see that if (1) converges, the derivatives of  $f(z)$  may be evaluated by differentiation under the integral sign. It is found, using (5), that

$$W^k[z^{\nu-1/2} f(z)] = z^\nu \left(\frac{2}{\pi}\right)^{1/2} \int_0^\infty t^{2k+1/2} K_\nu(zt) d\alpha(t).$$

If  $\alpha(t)$  is the indefinite integral of  $\varphi(t)$ , the change of variable  $t = uy$  leads to the result that

$$Q_k[f(u)] = I_k(u);$$

Theorem 1 now follows from the lemma.



The necessity of (3) for Theorem 3 is established similarly. To prove the sufficiency, let (3) be satisfied and consider

$$J_k(x) = \left(\frac{2x}{\pi}\right)^{1/2} \int_0^\infty t^{1/2} K_\nu(xt) Q_k[f(t)] dt.$$

The change of variable  $u = 2k/t$  reduces this to

$$J_k(x) = \frac{1}{(2k-1)!} \left(\frac{2kx}{\pi}\right)^{1/2} \int_0^\infty u^{-1} K_\nu(2kx/u) V_k[u^{-1/2} f(u)] du,$$

where

$$V_k[g(u)] = u^{2k-1} W_k[u^\nu g(u)]. \quad (6)$$

It is easily verified that  $V_k$  is an Euler operator.<sup>10</sup> If  $\bar{V}_k$  denotes the adjoint of  $V_k$ , we then have<sup>11</sup>

$$J_k(x) = \frac{1}{(2k-1)!} \left(\frac{2kx}{\pi}\right)^{1/2} \int_0^\infty u^{-1/2} f(u) \bar{V}_{k,u} [u^{-1} K_\nu(2kx/u)] du, \quad (7)$$

where  $\bar{V}_{k,u}$  operates with respect to  $u$ , provided that

$$u^{p+q+1} \frac{d^p}{du^p} [u^{-1/2} f(u)] \frac{\partial^q}{\partial u^q} [u^{-1} K_\nu(2kx/u)] \rightarrow 0 \quad (8)$$

as  $u \rightarrow 0$  and  $u \rightarrow \infty$ , for  $p, q = 0, 1, 2, \dots, 2k-1$ . Now the hypothesis of Theorem 3 implies in particular that  $g(u) = u^{-1/2} f(u)$  satisfies

$$g(u) = o(u^{-1/2}), \quad V_k[g(u)] = O(u^{-3/2}) \quad (u \rightarrow \infty).$$

By a theorem of the author<sup>12</sup> these relations imply that

$$g^{(p)}(u) = o(u^{-p-1/2}) \quad (u \rightarrow \infty; p = 1, 2, \dots, 2k-1).$$

The function  $h(u) = u^{-1} K_\nu(2kx/u)$  satisfies

$$h^{(q)}(u) = O(u^{-q-1/2}) \quad (u \rightarrow \infty),$$

as may be shown, for example, by another application of the same theorem, using (9) (below). Similar reasoning applies when  $u \rightarrow 0$ , and we obtain (8) and consequently (7). But  $u^{-1} K_\nu(2kx/u)$  is homogeneous of order  $-1$ , and hence<sup>13</sup>

$$\bar{V}_{k,u} [u^{-1} K_\nu(2kx/u)] = V_{k,x} [u^{-1} K_\nu(2kx/u)]. \quad (9)$$

We find, then, using (5) and (6), that

$$\begin{aligned} J_k(x) &= \frac{1}{(2k-1)!} \left(\frac{2kx}{\pi}\right)^{1/2} \int_0^\infty u^{-3/2} f(u) \left(\frac{2kx}{u}\right)^{2k} K_\nu\left(\frac{2kx}{u}\right) du \\ &= \left(\frac{2}{\pi}\right)^{1/2} \frac{(2k)^{2k+3/2}}{(2k)!} \int_0^\infty t^{2k+1/2} f\left(\frac{x}{t}\right) K_\nu(2kt) dt. \end{aligned}$$



By the lemma,

$$J_k(x) \rightarrow f(x) \quad (k \rightarrow \infty).$$

It is now easy to complete the proof along standard lines.<sup>14</sup>

<sup>1</sup> Watson, G. N., *A Treatise on the Theory of Bessel Functions*, 1922, p. 78.

<sup>2</sup> Meijer, C. S., *K. Akad. v. Wetenschappen te Amsterdam, Proc. Sect. Sci.*, **43**, 599-608, 702-711 (1940).

<sup>3</sup> Greenwood, R. E., *Ann. Math.*, **42** (2), 778-805 (especially 804) (1941).

<sup>4</sup> Widder, D. V., *Trans. Amer. Math. Soc.*, **39**, 244-298 (1936); further references are given in this paper.

<sup>5</sup> Other representation theorems are given by Meijer, op. cit., and by the present author in a paper forthcoming in *Bull. Amer. Math. Soc.*

<sup>6</sup> Widder, op. cit.

<sup>7</sup> Widder, D. V., *Trans. Amer. Math. Soc.*, **36**, 107-200 (especially 122 ff.) (1934).

<sup>8</sup> Meijer, op. cit., p. 601; Watson, op. cit., p. 219.

<sup>9</sup> Watson, op. cit., p. 79.

<sup>10</sup> Boas, R. P., and Widder, D. V., *Trans. Amer. Math. Soc.*, **45**, 1-72 (especially 34) (1939).

<sup>11</sup> *Ibid.*, **45**, 1-72 (especially 35) (1939).

<sup>12</sup> Boas, R. P., *Duke Math. Jour.*, **3**, 637-646 (especially 643) (1937).

<sup>13</sup> Boas and Widder, op. cit., p. 37.

<sup>14</sup> See, e.g., Boas and Widder, op. cit., p. 58.

## ON FORMAL EXPONENTIAL DIFFERENTIATION IN RINGS

By H. S. VANDIVER

DEPARTMENT OF PURE MATHEMATICS, THE UNIVERSITY OF TEXAS

Communicated November 21, 1941

The invariantive character of the operation

$$\frac{d^n(e^{av})}{dv^n} = a^n e^{av}$$

has been taken advantage of in many algebraic and analytical connections. Much of this work has been purely formal and algebraic, not really depending on the analytic properties of the exponential function. In order to make these formal methods as convenient as possible, in many applications, the writer sets up the concepts explained in the present paper. In particular, the notion of a polynomial in which the exponents may be members of an algebraic ring, not necessarily rational, is introduced.

Let  $R$  be a commutative ring including the ring of rational integers and write formally

$$A(x) = \alpha_1 x^{a_1} + \alpha_2 x^{a_2} \dots + \alpha_r x^{a_r} \quad (1)$$

where the  $\alpha$ 's are in  $R$  and *also the  $a$ 's, the latter not necessarily in the rational ring* and  $x$  is an indeterminate in  $R$  with  $x^0 = 1$  and  $1 \cdot x^a = x^a$ . We shall call (1) a formal polynomial in  $R$ . If we have also

$$B(x) = \beta_1 x^{b_1} + \beta_2 x^{b_2} \dots + \beta_t x^{b_t} \quad (2)$$

with now the  $a$ 's in (1) distinct as well as the  $b$ 's in (2), then

$$A(x) = B(x) \quad (3)$$

If and only if  $s = t$ , the  $b$ 's are the  $a$ 's in some order, and if  $a_i = b_j$  then  $\alpha_i = \beta_j$ . Also, for any  $A(x)$  and  $B(x)$  we have

$$A(x) + B(x) = B(x) + A(x). \quad (4)$$

For multiplication of these polynomials we use the rules

$$\alpha x^a \cdot \beta x^b = \alpha \beta x^{a+b} \quad (5)$$

$$\begin{aligned} (A(x) + \alpha x^a)B(x) &= B(x)(A(x) + \alpha x^a) \\ &= A(x)B(x) + \alpha x^a B(x). \end{aligned} \quad (6)$$

The use of the last two relations will enable us to express the product of any two polynomials as a polynomial, and similarly for any number of factors. Application of (6) also gives

$$\alpha_1 x^{a_1} + \alpha_2 x^{a_2} \dots + \alpha_n x^{a_n} = (\alpha_1 + \alpha_2 + \dots + \alpha_n) x^{a_1} \quad (7)$$

so that it follows that any formal polynomial may be reduced to a polynomial in which all the exponents of  $x$  appearing therein are distinct.

We shall now define *formal exponential differentiation* in this system. Write, using (1),

$$\bar{D}_x A(x) = \alpha_1 a_1 x^{a_1-1} + \alpha_2 a_2 x^{a_2-1} + \dots + \alpha_s a_s x^{a_s-1}. \quad (8)$$

This is said to be the *formal exponential derivative* of  $A(x)$  with respect to  $x$ . Note that if any of the  $a$ 's is zero the corresponding term vanishes in the derivative, corresponding to the ordinary notion that the derivative of a constant is zero. Repeated use of this formula gives the  *$n$ -th formal exponential derivative*,

$$\bar{D}_x^{(n)} A(x) = \alpha_1 a_1^n x^{a_1-n} + \dots + \alpha_s a_s^n x^{a_s-n}. \quad (9)$$

Henceforth we shall abbreviate the left-hand member of (9) as  $\bar{D}^{(n)}(A)$ , as differentiation will all be with respect to  $x$  in this paper and only polynomials in  $x$  will be employed.

We may prove easily by induction the relation

$$\bar{D}^{(n)}(AB) = \sum_{i=0}^n \binom{n}{i} \bar{D}^{(i)}(A) \bar{D}^{(n-i)}(B), \quad (10)$$

which is the analog of the Leibniz formula. More generally, by similar methods we find

$$\bar{D}^{(j)}(A_1 A_2 \dots A_s) = (A_1 + A_2 + \dots + A_s)^j, \quad (11)$$

where, in the expression on the right, we expand by the multinomial theorem and replace  $A_i^r$  by  $A_i^{(r)}$  with  $A_i^{(r)} = \bar{D}^{(r)}(A_i)$ ; the latter theorem is written in the form

$$(A_1 + A_2 + \dots + A_s)^j = \frac{j!}{c_1! c_2! \dots c_s!} A_1^{c_1} A_2^{c_2} \dots A_s^{c_s}, \quad (12)$$

the summation ranging independently over each set of positive or zero  $c$ 's satisfying

$$c_1 + c_2 + \dots + c_s = j,$$

and further  $A_i^0 = A_i$ .

We may introduce the quotient of two formal polynomials  $A/B$  where now all exponents are rational integers and  $R$  is a field, and handle it in a fashion similar to the way the quotient of two ordinary polynomials is treated. For the derivative of such an expression we make a definition so that if  $B \neq 0$  and

$$\frac{A}{B} = C$$

then

$$\bar{D}(A) = B\bar{D}(C) + C\bar{D}(B),$$

which gives

$$\bar{D}(C) = \frac{B\bar{D}(A) - A\bar{D}(B)}{B^2}, \quad (13)$$

after substituting  $A/B$  for  $C$  on the right.

Employing polynomials involving exponents which may not be rational, in the matter indicated above, we may prove the following theorem:

*Let  $R$  be the ring of algebraic integers in an algebraic field and put*

$$f_{n_i}^{(i)}(\alpha_1, \alpha_2, \dots, \alpha_{k_i}) = \sum_{r=1}^{k_i} \alpha_{ri} a_{ri}^{n_i}$$

*where the  $a$ 's and  $\alpha$ 's are in  $R$  and the  $n$ 's are rational integers  $\geq 0$ . Further let there be a rational integer  $d > 0$  such that for all  $r$ 's in the range 1 to  $k_i$  and all  $i$ 's in the range 1 to  $s$ ,*

$$a_{ri}^d \equiv 1 \pmod{\mathfrak{m}}$$

*where  $\mathfrak{m}$  is a fixed ideal in  $R$ . Also, let*

$$\beta_1 + \beta_2 + \dots + \beta_s \equiv 0 \pmod{\mathfrak{m}}$$

where the  $\beta$ 's are in  $R$  and suppose that  $\mathfrak{a}_i$  is the greatest common ideal divisor of the ideals

$$(a_{ri}), r = 1, 2, \dots, k_i,$$

each of which has a factor,  $\neq (1)$ , in common with  $\mathfrak{m}$ , then

$$(f_{n_1} + f_{n_2} + \dots + f_{n_s})^j \equiv 0 \pmod{(\mathfrak{m}^j, \mathfrak{a}_1, \mathfrak{a}_2, \dots, \mathfrak{a}_s)}$$

where we expand the left-hand member employing (12), and set

$$f_{n_i}^i = \beta^i f_{n_i}^{(i)} + id(\alpha_1, \alpha_2, \dots, \alpha_{k_i}), \\ i = 1, 2, \dots, s.$$

The proof of this follows rather closely the argument employed for the proof of a related theorem<sup>1</sup> given by the writer in another paper, after we introduce the generalized exponents referred to above.

The theorem of the present paper has many applications which I hope to consider elsewhere.

<sup>1</sup> *Bull. Amer. Math. Soc.*, **43**, 420–421 (1937).

---



PROCEEDINGS  
OF THE  
NATIONAL ACADEMY OF SCIENCES

Volume 28

February 15, 1942

Number 2

Copyright 1942 by the National Academy of Sciences

*RELATIVE EFFECTS OF X-RAYS AND NEUTRONS ON CHROMOSOMES IN DIFFERENT PARTS OF THE "RESTING STAGE"*

BY ALFRED MARSHAK\*

RADIATION LABORATORY, UNIVERSITY OF CALIFORNIA

Communicated December 1, 1941

*Introduction.*—When rapidly growing plant or animal tissues are treated with x-rays or with neutrons, it has been observed that at three hours after treatment the per cent anaphase cells showing no chromosome abnormalities is a negative exponential function of the dose.<sup>1</sup> It was demonstrated that these cells were at the onset of prophase at the time of irradiation. The ratio of the slope of the curve obtained with neutrons to that obtained with x-rays was approximately six for all species studied.<sup>2</sup> The constancy of this ratio was taken as evidence of the similarity of the structural and physiological conditions in the chromosomes of the different species at this stage. The investigations reported here show that within a single species the value of the neutron x-ray ratio varies for chromosomes irradiated in different parts of the resting stage and consequently that the conditions with respect to radiation response vary during this morphologically homogeneous stage of the nuclear cycle.

*Materials and Methods.*—The conditions for germination and irradiation with x-rays were the same as described in previous experiments and the same lot of seed of *Vicia faba* was used. For neutron studies the seedlings were arranged on a holder so that the root tips were in a collimated beam of neutrons produced by bombarding a beryllium target with 16 million volt deuterons with the 60-inch cyclotron.<sup>3</sup> The dose was measured in arbitrary "n" units as previously described.<sup>2</sup> I am indebted to Dr. P. C. Aebersold for measuring the neutron doses and to Miss Addi van Nouhuys for assistance in preparing slides and making counts. Root tip smears were prepared at 8, 12, 18 and 24 hours after irradiation. Only cells in mid anaphase and late anaphase, where the ends of the disjoining chromosomes were well separated from each other, were counted. Such cells were classified into two categories. Those showing no chromosome abnormalities

were called normal. Those showing one or more chromosome attachments or fragments were listed as abnormal.

*Results.*—The data obtained are tabulated below. Seven doses were given in duplicate and one in triplicate. These all agreed within less than 5 per cent with one exception, where the difference was about 7 per cent.

TABLE 1

TIME IN HOURS	X-RAYS				NEUTRONS			
	DOSE—r	NORMAL	ABNORMAL	% NORMAL	DOSE—"n"	NORMAL	ABNORMAL	% NORMAL
8	0	606	7	99.0	0	398	7	98.2
	50	579	131	81.5	5	1015	471	68.5
	50	270	81	77.0	10	907	614	59.6
	50	697	200	77.8	10	253	231	52.4
	102	63	63	50.0	15	188	276	40.6
	100	405	340	54.4	20	218	440	33.2
	150	88	161	35.5	25	66	238	21.7
	200	45	143	23.9	30	57	246	18.8
	250	6	40	13.1	30	31	106	22.6
	400	2	26	7.1	40	7	62	10.2
12	0	402	5	99.0	0	112	3	97.4
	55	1432	300	82.7	5	866	289	75.0
	50	215	39	84.6	10	662	579	53.2
	100	681	217	76.0	20	197	396	33.3
	200	292	270	52.0	20	168	371	31.2
	300	88	142	36.6	23	245	529	31.8
					29	94	300	23.8
					32	104	406	20.4
					44	24	114	17.4
					60	1	17	5.5
18	0	429	5	99.0	0	374	10	97.5
	50	362	47	88.6	5	470	141	77.0
	100	400	181	68.9	20	332	247	57.4
	122	438	210	67.6	20	298	190	61.1
	200	143	131	52.2	30	269	377	41.6
	300	111	190	36.9	40	210	428	33.0
	300	164	318	34.1	40	259	618	29.6
	400	60	171	26.0	50	157	473	24.9
	500	35	146	19.3	70	123	647	16.0
24	0	655	5	99.3	0	1214	14	99.0
	50	394	30	93.2	10	128	51	71.5
	122	475	142	77.0	20	768	740	51.0
	130	757	304	71.3	30	94	227	29.2
	200	304	162	65.2	30	53	115	31.6
	300	209	217	49.0	45	55	233	19.1
	400	63	106	37.2	61	25	212	10.6
	500	117	203	36.6				

In figure 1 the logarithm of the per cent normal anaphases is plotted against dose in roentgens and in figure 2 against "n" units of neutron ion-

ization. It is clear that at each time interval for x-rays and for neutrons the points are best fitted by a straight line through the origin.

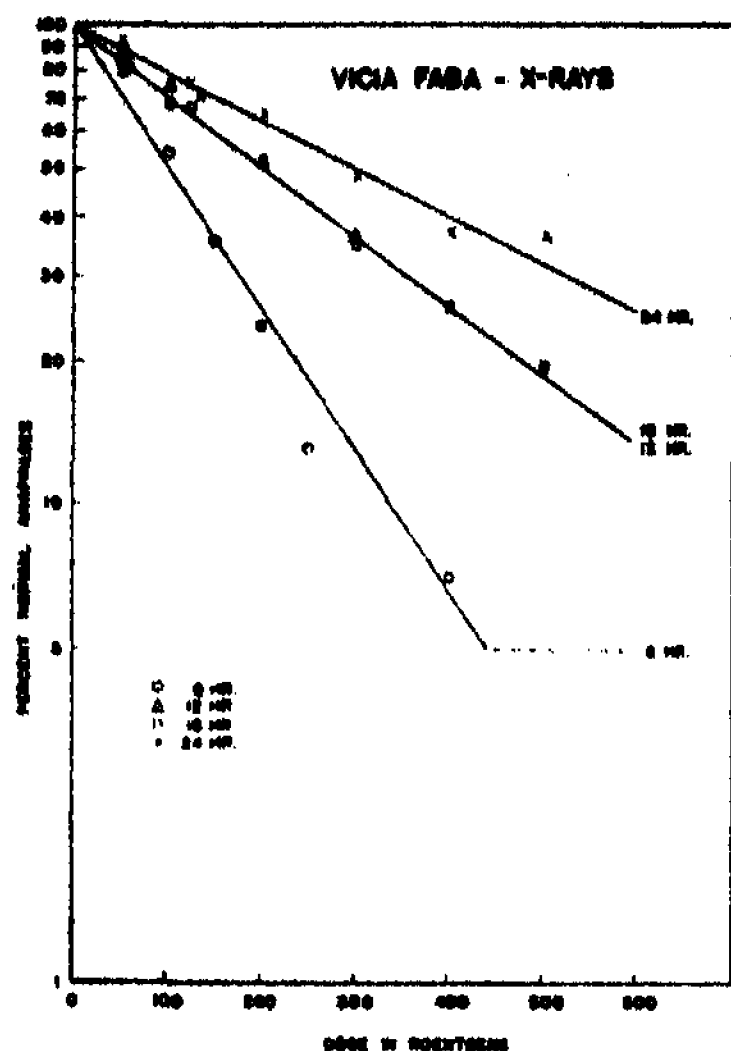


FIGURE 1

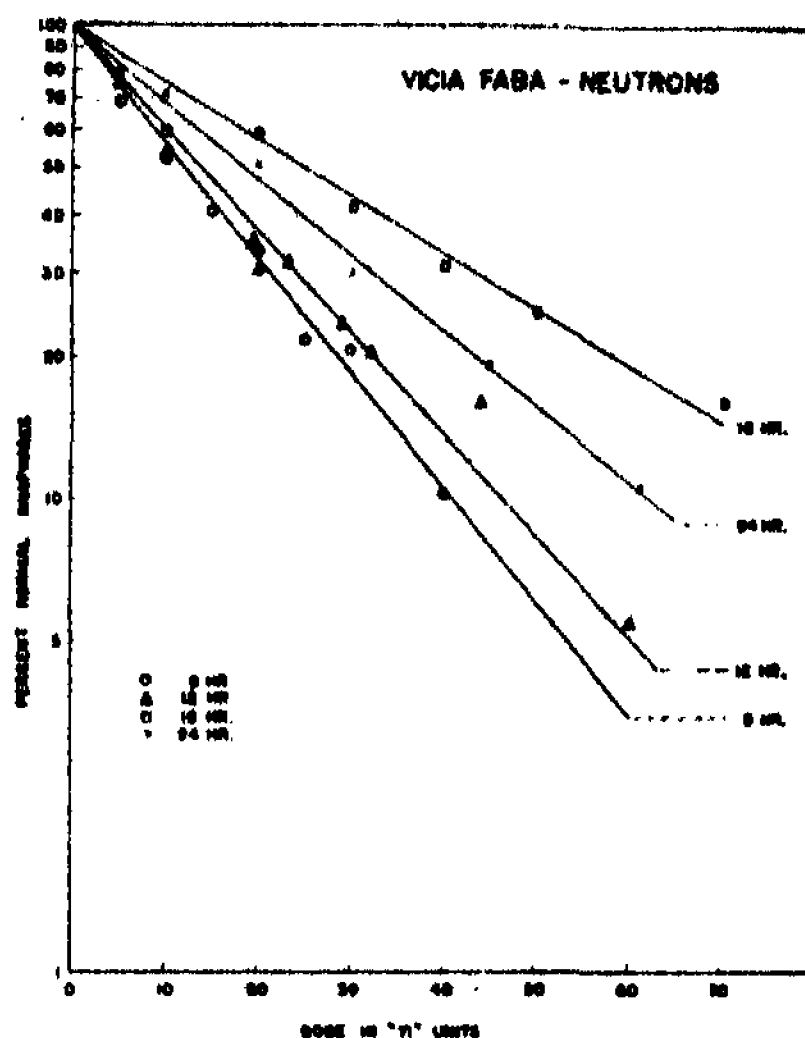


FIGURE 2

Fitting the lines graphically to the points, the slopes obtained are given in table 2. The values for three hours after irradiation are those which have previously been published.

TABLE 2  
SLOPES OF SURVIVAL CURVES

HOURS AFTER IRRADIATION	NEUTRONS	X-RAYS	NEUTRONS/X-RAYS
8	$7.0 \times 10^{-3}$	$1.1 \times 10^{-3}$	6.4
8	$5.6 \times 10^{-3}$	$6.9 \times 10^{-3}$	8.1
12	$5.0 \times 10^{-3}$	$3.3 \times 10^{-3}$	15.1
18	$2.8 \times 10^{-3}$	$3.3 \times 10^{-3}$	8.5
24	$3.7 \times 10^{-3}$	$2.2 \times 10^{-3}$	16.8

The observed negative exponential between the logarithm of the per cent *normal* anaphases and the dose means that the number of chromosome *abnormalities* produced is proportional to the dose when allowance is made for multiple hits in chromosomes of the same cell. As previously pointed out, the exponential survival curves indicate that a single effective agent produced by the x-rays or neutrons is sufficient for producing a chromosome abnormality.<sup>1, 2</sup> These results do not agree with those obtained by Sax, who found that chromosome aberrations in *Allium* and *Tradescantia*



increased as the 1.5 power of the x-ray dose.<sup>4, 5</sup> They do agree with those of Giles<sup>6</sup> for neutron effects on chromosomes of *Tradescantia* observed 30 hours after treatment, and with those of Creighton<sup>7</sup> for *Chorthippus* chromosomes 3 hours after irradiation with x-rays. These investigators did not examine the chromosome response as a function of dose in different parts of the nuclear resting stage.

It has been shown that the cells observed in anaphase 3 hours after irradiation were at the end of the resting stage at the time of irradiation.<sup>1</sup> This phase of the nuclear cycle was called the onset of prophase. One can also show that the chromosome abnormalities observed between 3 and 13 hours after irradiation were all in the resting stage when treated. This will be apparent from the following considerations.

Cells in anaphase 24 hours after irradiation show degenerating micronuclei. These arise originally from chromosome fragments which form separate small nuclei, the size depending on the number and size of fragments included. During the resting stage, chromonematic structure can clearly be discerned in the micronuclei as well as in the principal nucleus. However, during the prophase they become pycnotic and eventually diffuse in their staining properties. They may persist through the anaphase or lose their chromaticity to such an extent as to be barely distinguishable from the cytoplasm. Such degenerating micronuclei are not seen in anaphase as late as 18 hours after irradiation. It follows, therefore, that these cells were in the resting stage at the time of irradiation and the 24-hour cells were in some stage of the preceding nuclear cycle. (Since micronuclei must be attributed to abnormalities previously formed, they were not included in counts of chromosome abnormalities.) Comparison of the slopes of the neutron and x-ray curves ( $n/x$ ) for the interval 3-18 hours may then be used as an index of the relative sensitivity of the different portions of the resting stage to these agents. Such a comparison shows that the relative sensitivity to x-rays and neutrons varies by as much as a factor of two, and if the cells at 3 hours are included, by a factor of three.

One striking difference between the neutron and x-ray results is the steeper slope at 24 as compared with 18 hours in the case of neutrons but not with x-rays. Also at this time  $n/x$  is the highest observed (16.8). As explained above, the chromosomes in anaphase 24 hours after irradiation must have been in the preceding nuclear cycle when treated. The particular stage they were in is unknown, but since the time from the last resting stage corresponds to that for the interval from onset of prophase to anaphase, it may have been in the onset of prophase. The results therefore indicate that cells in which a chromosome abnormality has been produced by neutrons are much more likely to show another chromosome abnormality in the second succeeding anaphase than similar cells treated with x-rays. An adequate explanation for this observation may be found in a considera-

tion of the distribution of ionization of a proton track as compared with a beta-particle. If an abnormality is produced when the sensitive portion of a chromonema is transected by a proton which produces one ion pair in the sensitive volume, we may expect more energy to be released in the immediate vicinity, unless the diameter of this volume were much smaller than the average spacing of ion pairs along the track. This has been shown not to be the case (i.e., the diameter of the sensitive volume is of the same order of magnitude as the spacing of ions in a proton track,  $10^{-7}$  cm.). The distance between ion pairs along a beta-particle track is  $10^{-5}$  cm. and an effect by the second ion pair, therefore, is much less probable. Thus a cell showing abnormalities at 3 hours after irradiation is more likely to show abnormalities in the succeeding anaphase (21 or more hours later) than a cell with the same number of abnormalities produced by x-rays. Nishina and Moriwaki<sup>8</sup> have found unusually high frequencies of multiple mutations and chromosome aberrations in a single chromosome following neutron treatment of *Drosophila*, which is in accordance with the hypothesis given above.

As mentioned previously,  $n/x$  for chromosomes 3 hours after irradiation is consistently 6 in tissues as different as tomato root tips and mouse tumors. However, with the same chromosomes in different parts of the resting stage,  $n/x$  varies by a factor of at least two. Conditions within the chromosome must, therefore, be changing. Thus, in the nuclei which take 12 hours to reach anaphase, the intrachromosomal structures which respond to x-ray or neutron ionization differ from those of nuclei which take 3 hours. Previous experiments have shown that the sensitivity of the 3-hour nuclei to x-rays may be markedly altered by pH, but not in the 5 to 36-hour nuclei, indicating a difference in the mechanism of response of the two types of nuclei. It was suggested that the greater sensitivity to x-rays and the response to pH of the 3-hour nuclei was due to the presence of closely approximated charged surfaces formed by the division of the chromonemata at this time.<sup>9</sup> Accepting this hypothesis, the difference in relative susceptibility of 8, 12 and 18-hour nuclei remains unexplained. If, in order to produce a chromosome abnormality by irradiation of chromosomes in the resting stage, it is necessary for the ionization or excitation to break several chemical bonds, the probability of multiple breakage will be greater with neutrons than with x-rays.

At least two different physiological states in the "resting" nucleus may now be identified. One, 18 hours prior to anaphase, shows a sharp maximum in the formation of half-chromatid fragments which was attributed to a critical stage in the synthesis of new chromonemata.<sup>16</sup> The other is the 12-hour peak in neutron sensitivity described above. The greater efficiency of neutrons in the 12-hour nuclei where the synthesis is more complete may then be due to the presence of fewer of the less stable bonds pres-

ent at 18 hours and at 8 hours where the chromonemata are preparing for the actual separation which is observed 5 hours later.

Timoféeff-Ressovsky and Zimmer<sup>11</sup> found that neutrons were less efficient than x-rays in producing recessive lethal mutations in *Drosophila*. Dempster<sup>12</sup> has found that although neutrons are less efficient in producing recessive lethal mutations they are more efficient in producing dominant lethals and translocations. If production of a visible mutation requires alteration of only a single bond while a dominant lethal (deficiency) requires alteration of several as in the case of the microscopically observed chromosome abnormalities, the genetic results and the cytological data can both be accounted for by the hypothesis just advanced.

A theory for the mechanism of chromosome response to ionizing radiation at the onset of the mitotic prophase and the pachytene stage of meiosis has been proposed<sup>13</sup> and supported by subsequent investigations.† Another theory postulating an initial chromosome breakage followed by fusion or "healing" of broken ends has been applied to chromosome response at all mitotic stages.<sup>4, 5, 16</sup> The variation in the ratio  $n/x$  with the stage of mitosis indicates that the simple breakage-healing hypothesis is inadequate. It fails to explain why chromosomes taking 3 hours to reach anaphase are more sensitive to x-rays or neutrons than the thinner; and therefore, presumably more delicate ones which take 8 or 18 hours. It also fails to explain why chromonemata which take 12 hours to reach anaphase and are, therefore, less developed and thinner than those which take 8 hours should show a greater relative response to neutrons. If differential healing is invoked, it is difficult to explain why chromosomes taking 3, 8 and 18 hours to reach anaphase show relatively greater amounts of healing following neutron treatment than chromosomes which take 12 hours.

Since  $n/x$  increases during the resting stage, it follows that independently of the units used to measure x-rays and neutrons, the latter are capable of producing more abnormalities in resting nuclei than x-rays. This result is of interest in x-ray and neutron therapy, for it suggests that neutrons may produce regression of tumors resistant to x-rays. The amount of damage to skin in comparison with a given amount of tumor regression may be different from that observed with x-rays. The value of the "erythema dose" in terms of tumor response, therefore, needs to be independently determined in neutron therapy. Likewise, the daily "safe dose" in terms of equivalent ionization may be considerably lower with neutrons than with x-rays.

*Summary.*—1. The logarithm of the per cent normal anaphase chromosomes is a negative exponential function of the dose of neutrons or x-rays in different parts of the nuclear cycle. Chromosome abnormalities are (i.e., if allowance is made for multiple hits) therefore directly proportional to the dose at all stages studied.

2. The ratio of neutron to x-ray efficiency ( $n/x$ ) is constant for different species at the onset of prophase, but increases by a factor of at least 2 during the nuclear resting stage of a single species.

3. Increase in  $n/x$  during the resting stage indicates: (a) Conditions for chromosome response are not constant during resting stage. (b) The intrachromosomal structure responding to ionization may be different for the different parts of the nuclear cycle and resting stage. A theory invoking a single mechanism for the production of chromosome abnormalities is not adequate. (c) The relative effect of neutrons on chromosomes in the resting nucleus is greater than with x-rays. (d) Neutron therapy may produce regression in tumors which do not respond to x-rays. (e) The daily "safe dose" for exposure to neutrons may be considerably less than for x-rays.

4. The use of neutrons and x-rays as described furnishes a method for identifying the functional stages in the nucleus which cannot otherwise be identified.

5. The bearing of these results on theories concerning the mechanisms of the action of ionizing radiations is discussed.

\* Finney-Howell Research Fellow.

† See references 1, 2, 9, 10, 14 and 15 below.

<sup>1</sup> Marshak, A., *Proc. Nat. Acad. Sci.*, **23**, 362-369 (1937).

<sup>2</sup> Marshak, A., *Proc. Soc. Exp. Biol. Med.*, **41**, 176-180 (1939).

<sup>3</sup> Aebersold, P. C., *Phys. Rev.*, **56**, 714 (1939).

<sup>4</sup> Sax, K., *Genetics*, **25**, 41-68 (1940).

<sup>5</sup> Sax, K., *Ibid.*, **26**, 418-425 (1941).

<sup>6</sup> Giles, N., *Proc. Nat. Acad. Sci.*, **26**, 561-575 (1940).

<sup>7</sup> Creighton, M., *Exp. Zool.*, **87**, 347 (1941).

<sup>8</sup> Nishina, Y., and Moriwaki, D., *Sci. Papers of the Inst. of Phys. and Chem. Res.*, **38**, 371-376 (1941).

<sup>9</sup> Marshak, A., *Proc. Soc. Exp. Biol. Med.*, **39**, 194-198 (1938).

<sup>10</sup> Marshak, A., *Proc. Nat. Acad. Sci.*, **25**, 502-510 (1939).

<sup>11</sup> Timoféeff-Ressovsky, N. W., and Zimmer, K. G., *Naturwiss.*, **21**, 362 (1938).

<sup>12</sup> Dempster, E. R., *Proc. Nat. Acad. Sci.*, **27**, 249 (1941).

<sup>13</sup> Marshak, A., *Jour. Gen. Physiol.*, **19**, 179-198 (1935).

<sup>14</sup> Marshak, A., *Proc. Soc. Exp. Biol. Med.*, **38**, 705-713 (1938).

<sup>15</sup> Marshak, A., *Proc. Nat. Acad. Sci.*, **25**, 510-516 (1939).

<sup>16</sup> Gustafsson, A., *Hereditas*, **23**, 281-335 (1937).

## COMPARATIVE RATES OF DIVISION IN LARGE AND SMALL CELLS OF DEVELOPING FRUITS

By EDMUND W. SINNOTT

DEPARTMENT OF BOTANY, YALE UNIVERSITY

Communicated December 29, 1941

The relation of the size of a meristematic cell to its rate of division is an important problem in any analysis of growth and differentiation. A simple way to attack it is to determine the increase in number of cells in large-celled and in small-celled tissues of the same organ during a given period of growth. The early developmental stages of the cucurbit ovary provide good material for such a study. Its tissues are clearly delimited, and their volumes may thus be determined with some accuracy. Most of their constituent cells are relatively undifferentiated parenchyma. Cell volumes differ markedly in different tissues. By dividing tissue volume by cell volume, the approximate number of cells may be calculated for each tissue, and thus the comparative increase in cell number in large-celled and small-celled tissues, as growth proceeds, may be found.

In twelve inbred lines of cucurbits, belonging to four genera, tissue volumes and cell volumes were measured in ovaries with diameters of 2 mm. and in ovaries with diameters of 10 mm. In this stage of ovary development cell division is still going on throughout the organ. The tissues measured were the placental region, occupying the central portion of the ovary; the inner wall, extending from the placenta to the ring of main vascular bundles; the outer wall, extending from the bundle ring to the epidermis; and the epidermis itself, a single cell layer over the whole ovary. The ovary is essentially spherical and its structure was treated as a series of concentric spheres. Placental volume was calculated directly from placental diameter. Inner wall volume was found by subtracting this placental volume from the volume of a sphere of which the diameter was that of the bundle ring. The volume of the latter sphere subtracted from total ovary volume gave the volume of the outer wall. Average cell diameters (parenchymatous cells only) for these tissues had previously been reported in these same 12 lines.<sup>1</sup> Cell number has now been calculated for each tissue by dividing cell volume (cell diameter cubed) into tissue volume. In the epidermis, where cell division takes place only perpendicularly to the ovary surface, cell number was determined by dividing cell area (diameter squared) into area of ovary surface. These cell numbers are only approximations to the actual numbers but for purposes of comparison they are satisfactory and the consistent relationships which they show give confidence as to their essential accuracy.

There is a gradient of increasing cell volume from the outer tissues to

the inner ones. In the 2-mm. ovary the cells of the outer wall are about 2.2 as large as those of the epidermis, the cells of the inner wall 3.3 as large, and of the placenta 7.4 as large. In the 10-mm. ovary there has been a marked increase in cell volume except in the epidermis and the gradient is steeper, cells of the outer wall being about 5 times the volume of the epidermis, those of the inner wall 12 times, and those of the placental region over 40 times. These constant differences in cell volume between meristematic tissues show that the size to which a cell enlarges before it divides again is characteristic for a given tissue at a given stage of development. The problem is to determine whether this size is related to the rate of division.

In table 1 are given the average values, for the four tissues in the twelve lines, as to cell volume and number in the 2-mm. ovary, cell volume and number in the 10-mm. ovary, and ratio of increase in cell number (number at 10 mm. divided by number at 2 mm.). The latter value serves as a measure of comparative rate of cell multiplication in these tissues. The number of cell generations (and thus the number of divisions in each cell lineage) was found by determining the power of 2 which would give the observed ratio of increase.

TABLE 1  
RELATION OF CELL SIZE TO INCREASE IN CELL NUMBER

	OVARY DIAMETER		RATIO OF INCREASE IN CELL NUMBER	NUMBER OF CELL GENERATIONS
	2 MM.	10 MM.		
Epidermis				
Cell volume	1037 cu. $\mu$	1498 cu. $\mu$		
Cell number	139,346	3,115,670	22.90	4.52
Outer wall				
Cell volume	2362 cu. $\mu$	7636 cu. $\mu$		
Cell number	581,566	14,955,033	25.84	4.69
Inner wall				
Cell volume	3494 cu. $\mu$	17,724 cu. $\mu$		
Cell number	646,362	17,035,513	26.18	4.70
Placental region				
Cell volume	7713 cu. $\mu$	65,148 cu. $\mu$		
Cell number	133,670	3,324,341	25.00	4.64

It is evident that the ratio of increase in cell number, and thus presumably the rate of cell division, is essentially the same in all tissues and bears no relation to the initial size of the cells in the tissue or to the amount of increase in cell size as growth proceeds. That increase in cell number in the large-celled tissues is actually due to the division of these large cells rather than of small ones which might be scattered among them is indicated by the fact that the range in cell size in a tissue is not great, and that



mitotic figures and newly formed cell walls are frequent in these large cells. The author<sup>2</sup> has recently shown that relatively large, vacuolate cells like those here discussed commonly divide in meristematic regions.

For this material, at least, there is no relation between the volume of a cell and the rate at which it divides. In larger cells, the amount and rate of cell expansion, from the time when a new daughter cell is formed until this cell is ready to divide again, is evidently greater than in smaller cells. Whether this increase is in protoplasm or only in the size of the vacuole is not known.

Rate of cell division seems to be determined by some factor independent of cell size and operative throughout the entire organ.

<sup>1</sup> Sinnott, E. W., *Amer. Jour. Bot.*, 26, 179 (1939).

<sup>2</sup> Sinnott, E. W., and Bloch, R., *Ibid.*, 28, 225 (1941).

---

## CHROMOSOME DEGENERATION IN RELATION TO GROWTH AND HYBRID VIGOR

BY D. F. JONES

CONNECTICUT AGRICULTURAL EXPERIMENT STATION

Communicated January 3, 1942

In the twenty-five years of genetic research, since hybrid vigor was put on a Mendelian basis,<sup>9</sup> additional evidence has accumulated from many sources that contributes toward a further solution of this problem. While much of this new information confirms the original hypothesis an important situation is brought to light which was not fully visualized in the early days of genetic investigation and shows where additional information is needed both for a more complete understanding of the principles involved and for the best application of hybrid vigor to practical breeding.

Both Shull and East showed clearly that the greatest vigor, generally, is exhibited when germinal heterogeneity is at the maximum. When this diversity is reduced by Mendelian recombination vigor is lost. The phrase "stimulus of heterozygosis" was contracted to the word "heterosis" by G. H. Shull<sup>23</sup> (1914) and was considered by Shull, East and other investigators at this time to be something additional to the normal expression of hereditary characters. They assumed that it could not be fixed and would therefore never be exhibited by organisms when in the homozygous condition. An interaction of diverse nuclear elements within an unfamiliar cytoplasm was postulated by A. F. Shull<sup>22</sup> (1912) as a possible contributing factor, based on evidence from rotifers. A revised form of the original hypothesis was put forth by East<sup>6</sup> (1936) which assumed an interaction between diver-

gent alleles. Evidence for this he found in segregating progenies of *Nicotiana* where few gene differences were involved. There was marked skewness in the frequency distributions of size characters but the extreme plus variants were impossible to fix whereas less difficulty was encountered in stabilizing minus variants. Allelic differences in the heterozygous condition in *Nicotiana* have shown a variation from each parental type that under certain conditions might be advantageous<sup>10</sup> (Jones, 1921). Dunn<sup>4</sup> (1937) describes lethal alleles in mice that interact to give viable heterozygotes. In *Godetia*, Hiorth<sup>8</sup> (1940) has found that alleles have an additive effect as postulated by East. But the evidence, so far available, does not indicate that this type of interaction has an important part in the production of hybrid vigor.

All stimulation hypotheses, based on heterozygosis, ignore the fact that many plants that are almost completely self-fertilized in every generation are still the most efficient producers in the list of cultivated plants and many closely bred wild species of both plants and animals show no lack of ability to main themselves under natural conditions. Several cases are on record where recombination produces homozygous individuals that are equal or nearly equal to the hybrids from which they came as Malinowski<sup>16</sup> (1935) has shown in *Phaseolus* and Sveschnikova<sup>25</sup> (1940) in *Vicia*. Dodge<sup>3</sup> (1942) has obtained a recombination of growth-promoting genes in haploid *Neurospora* where heterozygosis is not involved. By back-crossing maize hybrids to one parent in successive generations and then crossing by the other, Richey and Sprague<sup>19</sup> (1931) have evidence for an increased or equal growth with reduced heterozygosity in the same gene complex. On the other hand Randolph<sup>18</sup> (1942) has found that tetraploid maize heterozygotes are relatively more vigorous than tetraploid homozygotes and this may be good evidence for an interaction between dissimilar genes. But it may also be simply an abnormal situation in which the weaker homozygotes are at a disadvantage while the much more vigorous heterozygotes are not.

From all the evidence at hand homozygous organisms are as well able to live as heterozygous forms and there is no necessity to assume any inherent physiological advantage in germinal association of dissimilar elements, in itself, for which the term, heterosis, was originally devised. The word continues to be used and has become practically synonymous with hybrid vigor. A third of a century of biological research has failed to bring forth any clear evidence in support of the stimulation theory. On the other hand, there is much new evidence to show that dominance and an accumulation of favorable heredity are the principal factors involved.

When naturally cross-fertilized organisms are inbred many defective and abnormal individuals are brought to light as is well known. Natural selection removes most of these and in any breeding program they are always selected away from unless valued for some particular purpose. Inbred



progenies after these defective and subnormal individuals are eliminated are still much reduced in size and in ability to grow and to reproduce. This is particularly true in maize, many other naturally cross-fertilized plants and nearly all domesticated animals. When unrelated inbred families are interbred there is an immediate rise in the growth level. The elimination or suppression of these visible abnormalities apparently has little to do with this phenomenon. This has led East<sup>6</sup> (1936) to say that "the elimination of deleterious recessives is of little importance in practical breeding and of no consequence whatever in the solution of the problem (of heterosis)."

This statement may be true for the defective genes that have a visible effect and can be eliminated but it is certainly not true for the non-visible defectives that are wide-spread in all cross-fertilized organisms and cannot be eliminated entirely by any method of selection now known. Heretofore their number and importance have not been appreciated. They can be detected only in favorable material when studied in such a way as to separate their action from that of all the remaining inherited complex, as recent investigations show clearly.

Many deficient loci in the corn chromosomes remain after long-continued inbreeding has eliminated the visible defectives. They exist in the homozygous condition but do not produce any distinct effect and are not as yet classified as Mendelian genes. They are similar in action to the visible defectives in that they reduce or slow physiological activity. In the long-continued self-fertilized lines of corn grown at the Connecticut Agricultural Experiment Station for more than 30 years, changes have taken place that delay flowering of certain progenies without otherwise altering their size. Other transmissible alterations reduce the size of the ear and height of the plant. None of these segregate clearly and the effects of any one are so small that they have been overlooked up to the present time.

Radiation experiments with many organisms show that chromosomes are frequently made deficient in the ability to transmit normal growth. Beadle and Tatum<sup>1</sup> (1941) irradiated the fungus, *Neurospora*, and found strains that lack the ability to produce certain growth-promoting substances. When chromosomes are broken and rearranged there is frequently a reduction of normal efficiency accompanying the break. While the evidence is far from complete it seems probable that many breaks, both natural and induced, reunite in the original arrangement but with various degrees of deficiency or derangement at the place of rupture. Whether the alteration precedes or follows the break is not important for the problem at hand. In *Drosophila*, collected from the wild, Dobzhansky and Queal<sup>2</sup> (1938) determined that about 39 per cent of the third chromosomes have loci that reduce viability below normal as compared with 2 per cent that raise it, and 3.5 per cent with visible external effects. Spencer<sup>24</sup> (1942) calculates that one detectable mutant is present in the heterozygous form

in about every four flies tested, varying somewhat in different species. This situation is probably the same, to a greater or less degree, in all naturally cross-fertilized organisms.

In naturally self-fertilized plants any hybrid vigor effect must be added to an already high level of productiveness. In plants like maize that are markedly reduced by inbreeding, the invigorating results of crossing are much more apparent. The extensive program of corn improvement by crossing inbred strains which has given such notable practical results in recent years has also brought forward evidence of much theoretical import. Well over 25,000 inbred strains of corn have been produced. Out of these, several hundred have been selected and are being used in the commercial production of hybrid corn in all parts of the country. The production of even the best of these inbreds does not exceed, and seldom approaches, 50 per cent of the yield of the original variety from which they were derived.<sup>13</sup> (Lindstrom, 1939). The few inbreds which are most widely used are not the most productive. In fact, some are so weak and unproductive that inbred seed cannot be produced regularly in those sections where their hybrids are most commonly grown. Many investigations have shown clearly that the ability to transmit yield is not closely correlated with characters visible in the inbreds.

Any two of these selected inbreds, if they are unrelated, can be crossed to give a marked increase in size of plant and yield of grain. Not all combinations have been tested but a sufficiently large sample has been tried to show that heterosis is practically universal among them. Some combinations are much better than others. In fact, the inbreds themselves are largely selected for their ability to give valuable hybrids. All of the hybrids are either within or close to the normal range of 80 to 120 per cent of the yield of the original varieties from which the inbreds were derived, well out of the 20 to 40 per cent range that the inbreds themselves are in. Any one of these—*A*, *B*, *C* and so on—can be crossed to give a vigorous hybrid. Since inbred *B* combines with inbred *C* to give as good a hybrid as *A* and *B* or *A* and *C* then *B* and *C* must differ from each other as much as *A* does from *B* or *C*. Similarly it is evident that *A* differs from *D*, *E* and so on down the list of hundreds of inbreds. No limit is in sight to the number of homozygous inbreds all of which differ from each other in whatever it is that makes up the major heterotic effect.

Not only do all combinations of two inbreds behave in this way but also all combinations of three and four or more. Just as in the single crosses, having two inbred parents, many double crosses, with four inbred parents, are better than others using the same inbreds but these differences are of an entirely different order of magnitude than the differences between inbreds and hybrids. In their range of variation in yield, inbreds and hybrids of the same maturity season seldom meet and even when they overlap the

individuals of each group clearly fall into two distinct frequency distributions<sup>11</sup> (Jones, 1922). The coefficients of variability for weight of grain per plant, which sums up the ability of the plant to grow and is a measure of reproductive ability, are seldom significantly more for double crosses than for single crosses. The importance of this fact has not been fully recognized.

In the combination of  $A B \times C D$ , bringing together four inbreds by successive crosses, every combination of chromosomes from none of  $A$  and all of  $B$  to all of  $A$ , and none of  $B$  unite in fertilization with ten chromosomes of  $C$  and  $D$  origin similarly assorted. Not only are the chromosomes shuffled as units but they are also taken apart and reassembled piece by piece down to the smallest amounts of each chromosome that can be separated by normal crossing over. In the 32,000,000 acres of double-crossed corn grown in the United States in 1941, a large sample of every possible combination of chromosome parts from the original inbreds has been taken. The astounding fact that practically every one of these plants, when given suitable conditions in which to grow, is normally vigorous and productive, proves beyond doubt that every seriously defective locus in any part of the chromosome complex is prevented from having an appreciable effect. This extraordinary situation points clearly to the fact that there are an enormous number of loci where loss or derangement may occur.

After inbreeding, the lethals and the most serious defectives are automatically eliminated. The resulting inbred strains, as far as they can survive, are reduced to the inbred level of vigor. Their position within the range of variation, possible at this level, depends upon the number and nature of their invisibly defective loci, to a large extent, irrespective of the potentially good heredity they may carry.

When unrelated inbreds are crossed practically all defective loci are covered by normal alleles. Complete or nearly complete dominance is shown. Visibly defective genes have been tested and shown to have little or no effect on yield in the heterozygous condition.<sup>12,16</sup> (Jones and Singleton 1935; Mangelsdorf, 1926). If differences between heterozygous and homozygous normals do exist they are small and of little consequence. With their defective loci covered the heterozygotes are lifted from the inbred to the crossbred level of vigor. As long as these defects are kept covered, crossbred individuals are restored to a condition that is comparable to that shown by naturally self-fertilized plants. In this condition their normal heredity has a chance to operate. On this plane of normal growth the inheritance determines height of plant, diameter of stalk, time of flowering and maturity, heat and cold tolerance, resistance to insects and diseases, and so forth. In other words, the plants respond to the interaction of all those qualitative and quantitative characters that are governed by the many complementary, duplicate, cumulative and inhibiting genes

that have been studied in many plants and animals. Some of these show dominance and some do not. Genes that reduce or inhibit large size and most rapid growth are known. All of these working together determine the position of the hybrid individual within the range of variation at the normal level.

Many genes promoting the most efficient growth and reproductive ability are at least partially dominant. Robbins<sup>20,21</sup> (1940, 1941) has compared plant hybrids with their parental types in tissue cultures. The offspring have the ability of both parents to produce and to use growth-promoting substances, some of which are lacking in one or the other parent. Many instances of this pooling of hereditary resources are well known.

The inheritance of many quantitative characters has necessitated the assumption of an absence of dominance. To many geneticists it has seemed unreasonable to use dominance to account for one phenomenon and to assume a failure of dominance in a different situation. But the facts are clear that many individuals occupy an intermediate position when compared with their parents in many measurable characters, particularly those that are little influenced by the environment. East<sup>5</sup> (1916) crossed *Nicotiana* species with flowers differing in length. The first generation was close to the arithmetical mean of the two parents in this linear character. Flower size is little affected by the size of the plants or the environmental conditions in which the plants grow. But in this same material, heterosis is clearly shown in height of plant and in the number of flowers and seeds.

MacArthur<sup>14</sup> (1941) found tomato hybrids to be almost exactly intermediate in fruit size when measured by the geometric mean. Here the character measured is three dimensional. Powers<sup>17</sup> (1941) also studied a wide species cross in tomatoes and found partial dominance of small fruit size when measured by the arithmetic mean, also dominance of smaller number of fruits per unit of vine. But many tomato crosses show a large increase in total amount of plant growth and in early maturity of fruit. Hayes and Jones<sup>7</sup> (1917) found that some varietal crosses of tomatoes gave an increased number and average weight of fruit as well as earlier maturity while other crosses did not.

Naturally self-pollinated plants show heterosis which is lost in later generations. But when hybrid corn is allowed to interpollinate, or is again inbred, not only is there a reduction in the number of normal, favorable growth genes but also there is the uncovering of the many defective loci always present in every part of the chromosome structure. The plants not only segregate within the normal complex but are rapidly reduced to the level of vigor in which the homozygous inbreds operate. Here normal heredity has a limited opportunity to find visible expression. This is proved by the fact that many of the most valuable inbred strains of corn are themselves weak and unproductive, as mentioned previously. Since

they combine well with nearly all other strains they undoubtedly have good heredity but this is prevented from expressing itself because something essential is missing. All other inbreds apparently have what these good combiners lack. The fact that any part of the chromosome structure of one inbred can substitute for the homologous part from almost any other inbred shows that the mechanism must be fundamentally alike within the species. Only comparatively minor derangements keep the chromosomes from functioning at full efficiency. This renders improbable the necessity of assuming extensive differences in germinal construction. Otherwise many combinations would fail to give normal growth.

For the practical breeder the maintenance of homozygous inbred lines is a matter of serious concern. The possibility of continued production of degenerative loci must be guarded against by progeny testing or valuable lines will be lost or reduced to such a low level of vigor that they can be used only with difficulty. For theoretical genetics it is worth while to note that there is the possibility of continued degeneration. This takes place in crossbred as well as in inbred families and has nothing to do with inbreeding but is more apparent after consanguineous mating and more quickly eliminated. The prevention of chromosome degeneration and the elimination of defective loci after they are formed, in organisms not exposed to rigorous natural selection, is one of the major biological problems of the future.

- <sup>1</sup> Beadle, G. W., and Tatum, E. L., *Proc. Nat. Acad. Sci.*, **27**, 499-506 (1941).
- <sup>2</sup> Dobzhansky, T., and Queal, M. L., *Genetics*, **23**, 463-484 (1938).
- <sup>3</sup> Dodge, B. O., *Ibid.*, **27**, 140-141 (1942).
- <sup>4</sup> Dunn, L. C., *Proc. Nat. Acad. Sci.*, **23**, 474-477 (1937).
- <sup>5</sup> East, E. M., *Genetics*, **1**, 311-333 (1916).
- <sup>6</sup> East, E. M., *Ibid.*, **21**, 375-397 (1936).
- <sup>7</sup> Hayes, H. K., and Jones, D. F., *Conn. Agric. Exper. Sta. Rept.* (1917), of 1916, Part V.
- <sup>8</sup> Hiorth, G., *Hereditas*, **26**, 441-453 (1940).
- <sup>9</sup> Jones, D. F., *Genetics*, **2**, 466-479 (1917).
- <sup>10</sup> Jones, D. F., *Ibid.*, **6**, 433-444 (1921).
- <sup>11</sup> Jones, D. F., *Amer. Nat.*, **56**, 166-173 (1922).
- <sup>12</sup> Jones, D. F., and Singleton, W. R., *Conn. Agric. Exper. Sta. Bull.*, **376** (1935).
- <sup>13</sup> Lindstrom, E. W., *Proc. Seventh Internat. Genet. Congr.*, pp. 191-196 (1939).
- <sup>14</sup> MacArthur, J. W., *Jour. Hered.*, **32**, 291-295 (1941).
- <sup>15</sup> Malinowski, E., *Zeit. ind. Abst. Vererb.*, **70**, 96-124 (1935).
- <sup>16</sup> Mangelsdorf, P. C., *Conn. Agric. Exper. Sta. Bull.*, **279** (1926).
- <sup>17</sup> Powers, L., *Jour. Agric. Res.*, **63**, 149-174 (1941).
- <sup>18</sup> Randolph, L. F., *Genetics*, **27**, 163 (1942).
- <sup>19</sup> Richey, F. D., and Sprague, G. F., *U. S. D. A. Tech. Bull.* **267** (1931).
- <sup>20</sup> Robbins, W. J., *Bull. Torrey Bot. Club*, **67**, 565-574 (1940).
- <sup>21</sup> Robbins, W. J., *Amer. Jour. Bot.*, **28**, 216-225 (1941).
- <sup>22</sup> Shull, A. F., *Biol. Bull.*, **24**, 1-13 (1912).
- <sup>23</sup> Shull, G. H., *Zeit. ind. Abst. Vererb.*, **12**, 97-149 (1914).
- <sup>24</sup> Spencer, W. P., *Genetics*, **27**, 171 (1942).
- <sup>25</sup> Sveschnikova, I. N., *Jour. Hered.*, **31**, 349-360 (1940).

TEMPERATURE AND "SEX-RATIO" IN *DROSOPHILA PSEUDOÖBSCURA*

BY C. D. DARLINGTON AND TH. DOBZHANSKY\*

JOHN INNES HORTICULTURAL INSTITUTION, LONDON, AND DEPARTMENT OF ZOÖLOGY,  
COLUMBIA UNIVERSITY, NEW YORK

Communicated January 5, 1942

The "sex-ratio" condition was found originally in some natural populations of *Drosophila affinis*<sup>1</sup> and subsequently in *D. obscura*,<sup>2</sup> *D. pseudoöbscura*<sup>3</sup> and certain other species. The offspring of a "sex-ratio" male consists, regardless of the genetic constitution of his mates, of daughters and few or no sons.

"Sex-ratio" occurs, so far as we know, only in species with two-armed *X*-chromosomes, and it is always inherited as though it were due to a gene located in the right (longer) limb of this chromosome. Any "sex-ratio" *X*-chromosome in race A of *D. pseudoöbscura* carries, however, three inversions in its right limb.<sup>4</sup> The nature of the association between the "sex-ratio" and these inversions is unknown. Nevertheless, the proximate mechanism of its action is reasonably clear; namely, in most of the first spermatocytes of "sex-ratio" males the *X*-chromosome is clearly quadripartite instead of bipartite; this chromosome undergoes two equational divisions, and all the resulting spermatids carry an *X*-chromosome. The *Y*-chromosome in the spermatocytes is strongly pycnotic, its centromere is inactive during the divisions, and the chromosome is eventually excluded from the nucleus and lost in the cytoplasm.<sup>5</sup> There is, however, a certain amount of variability in the spermatogenesis of "sex-ratio" males. Some spermatozoa come to carry an *X* and a *Y*, others only a *Y*, and still others neither an *X* nor a *Y*. The two latter classes give rise to the few sons that occur in the progeny of "sex-ratio" fathers. The *XO* sons are, of course, sterile.

Now it is well known that the sex chromosomes of animals frequently differ from the autosomes in their behavior at meiosis in the *XY* sex. They differ in two respects. The first of these is in the nucleic acid charge, which is frequently excessive in prophase and metaphase, as well as in the resting stage. The second is in the timing of the division cycle of the gene string and of the centromere; in the Heteroptera, for example, both of these seem to be advanced to give a coördinated suppression of pairing and crossing over and a division one mitosis ahead of the autosomes.<sup>6</sup> In *Drosophila* the excess charge is slight and the difference chiefly expresses itself in the mode of pairing and crossing over.<sup>6</sup>

The association of an increase in nucleic acid charge and hastening of division cycles we now know to be expected on more general grounds.<sup>7</sup> It



is to be inferred equally in "sex-ratio" *Drosophila*. *X* is supercharged enough to cause super-division. *Y* is presumably so much supercharged that its super-division cannot be seen. It is immobilized. The excessive nucleic acid charge goes, as it so often does, with a persistence of the nucleolus to metaphase.<sup>8</sup> This behavior, however, is slightly variable and might therefore provide a first means of testing the mode of action of a nucleic acid charge by experiment in the following way. In various plants

TABLE I  
NUMBER OF OFFSPRING AND PER CENT MALES IN "SEX-RATIO" CULTURES AT DIFFERENT TEMPERATURES

$^{\circ}$	ARRIES	NUMBER OF CULTURES	INDIVIDUALS COUNTED	PER CENT MALES
25°	I	5	485	4.80
	II	5	663	6.03
	III	7	665	7.22
	IV	8	1263	6.02
	V	6	403	5.71
	VI	5	736	6.79
Total		36	4188	6.18
22°	I	7	2516	2.07
	II	8	2924	2.74
	III	14	3648	3.56
	IV	9	2309	2.99
	V	4	1066	3.75
	VI	12	3135	6.09
Total		54	15,598	3.74
16.5°	I	7	2102	0.67
	II	5	1994	1.05
	III	10	3969	1.18
	IV	7	2030	1.08
	V	8	2557	0.74
	VI	9	3010	2.19
Total		46	15,662	1.22

and animals which can be grown satisfactorily over a temperature range of 0°C.-25°C., it has been possible to show that the lower extreme of temperature reduces the nucleic acid charge at metaphase of mitosis and meiosis.<sup>9,10</sup> Such a range of temperature is too much to ask of the fly. But with the range that is possible, it would appear worth while to look for a change in "sex-ratio" behavior.

For the purposes of the experiment, a "sex-ratio" strain originally derived from the population of Andreas Canyon, Mount San Jacinto, Cali-

fornia, was used. This strain was perpetuated with the aid of the sex-linked recessives yellow, singed, vermillion, compressed and short ( $y\ sn\ v\ co\ sh$ ) which served as markers. "Sex-ratio" is almost completely linked to  $co$ . Females of the constitution "sex-ratio"/ $y\ sn\ v\ co\ sh$  are normal wild type in appearance and breeding habit; they are crossed to  $y\ sn\ v\ co\ sh$  males in each generation.

A group of about six "sex-ratio"/ $y, sn\ v\ co\ sh$  females and an equal number of  $y\ sn\ v\ co\ sh$  males were transferred daily to fresh culture bottles. The bottles with the eggs deposited in them were divided into three groups. One group developed in an incubator at  $25.0 \pm 0.1^\circ\text{C}$ . The second group was left in a room in which the temperature fluctuated irregularly from about  $20^\circ$  to  $24^\circ\text{C}$ ., the average being close to  $22^\circ\text{C}$ . The third group developed in a cold room with an average temperature of  $16.5^\circ$  and four to five hourly fluctuations from  $15^\circ$  to  $17^\circ\text{C}$ . The wild type males hatching in these cultures must carry "sex-ratio." Five to six such males of each group were outcrossed to an equal number of females from a wild strain originating at Keen Camp, San Jacinto, California, which had developed in the cold room. The parents were transferred to fresh culture bottles daily or at two-day intervals. Oviposition took place at the same temperature at which the males had developed, but later all the cultures were kept together at room temperature. Complete counts of the offspring were made. We are indebted to Mr. B. Spassky for assistance in making these counts. A summary of the data is presented in table 1.

It is evident that "sex-ratio" is most extreme at the lowest temperature tried,  $16.5^\circ\text{C}$ ., at which only  $1.22 \pm 0.09$  per cent of the progeny are males. At room temperature the proportion of males rises to  $3.74 \pm 0.16$ , and at  $25^\circ$  to  $6.18 \pm 0.37$  per cent. The differences are undoubtedly significant. Although all the experimental cultures contained flies of similar pedigree, we have recorded separately the series of cultures coming from the same group of parents. As shown in table 1, six such series have been raised at each of the three temperatures, the number of cultures per series ranging from 4 to 14. A study of variance discloses a significant non-homogeneity of the results in the different series, especially those kept at room temperature. Since, as stated above, the "room temperature" varied rather widely and irregularly, the observed non-homogeneity presumably indicates a great sensitivity of "sex-ratio" to temperature changes.

Thus the lower temperature seems to increase the frequency of X-sperm produced by "sex-ratio" males from 94 to 99 per cent. How this result comes about is another matter. The difference is too slight to make a cytological discrimination feasible. It may be assumed that any effect on the Y is irrelevant, and that it is only the regularity in double division of the X that is changed as temperature changes. Since this was the basis of our expectation it is to differences in the nucleic acid charge that we



must look for the effective agent in this change. In this respect the effect of temperature is the opposite to that found in Trillium and Triton. That in itself is not surprising, for in them the metaphase charge is concerned while in *Drosophila* it is the resting stage reproduction. And between resting stage and metaphase there is a reversal of charge in Trillium and Triton. Thus the two parts of the relationship: temperature—nucleic acid charge—division cycle, are seen pieced together, although how they are pieced together we do not yet know.

*Summary.*—The “sex-ratio” gene or complex in the *X*-chromosome of many *Drosophila* species causes the *X*-chromosome to divide twice at meiosis in the males while the *Y* is thrown out. This abnormality seems to depend on an excessive nucleic acid charge and should therefore (on other evidence) be affected by temperature. This expectation was confirmed. The proportion of *X*-sperm was reduced from 99 to 94 per cent by raising the temperature from 16° to 25°C.

\* Experimental data by Th. Dobzhansky, the general interpretation by C. D. Darlington.

<sup>1</sup> Morgan, T. H., Bridges, C. B., and Sturtevant, A. H., *Bibliogr. Genetica*, 2, 1-262 (1925).

<sup>2</sup> Gershenson, S., *Genetics*, 13, 488-507 (1928).

<sup>3</sup> Sturtevant, A. H., and Dobzhansky, Th., *Ibid.*, 21, 467-490 (1936).

<sup>4</sup> Dobzhansky, Th., *Scientia Genetica*, 1, 67-75 (1939).

<sup>5</sup> Darlington, C. D., *Jour. Genetics*, 39, 101-137 (1939).

<sup>6</sup> Darlington, C. D., *Amer. Naturalist*, 68, 374-377 (1934).

<sup>7</sup> Callan, H. G., *Jour. Genetics*, 41, 349-374 (1941).

<sup>8</sup> Darlington, C. D., and Thomas, P. T., *Proc. Roy. Soc.* (in press).

<sup>9</sup> Darlington, C. D., and La Cour, L., *Jour. Genetics*, 40, 185-213 (1940).

<sup>10</sup> Callan, H. G., *Proc. Roy. Soc.* (in press).

## A GENERAL THEOREM ON THE INITIAL CURVATURE OF DYNAMICAL TRAJECTORIES\*

BY EDWARD KASNER AND DON MITTLEMAN

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Communicated December 24, 1941

The theorem we intend to establish is an extension to acceleration fields of higher order of Kasner's dynamical theorem<sup>1</sup> which states that: *If a particle starts from rest in any positional field of force, the initial curvature of the trajectory is one-third of the curvature of the line of force through the initial position.*

We formulate this result analytically. If  $\phi(x, y)$  and  $\Psi(x, y)$  represent the rectangular components of a planar field of force acting at any point

$(x, y)$  and  $t$  is the time, then the newtonian equations of motion for a particle of unit mass are

$$\frac{d^2x}{dt^2} = \phi(x, y), \quad \frac{d^2y}{dt^2} = \Psi(x, y).$$

If the curvature of the lines of force is everywhere zero, then the lines of force are straight lines, and a particle starting from rest at any point must necessarily move along the line of force. Thus, the trajectory and the line of force coincide and the theorem is trivially true. Excluding this degenerate possibility, we choose a point  $(x_0, y_0)$  at which the curvature of the line of force, which will be represented by  $\Omega$ , is not zero. By direct calculation, we find that the curvature of the line of force, which line of force satisfies the differential equation  $\frac{dy}{dx} = \frac{\Psi(x, y)}{\phi(x, y)}$ , is

$$\Omega = \frac{\phi(\Psi_x\phi + \Psi_y\Psi) - \Psi(\phi_x\phi + \phi_y\Psi)}{(\phi^2 + \Psi^2)^{3/2}}, \quad (1)$$

where the subscripts  $x$  and  $y$  denote partial differentiation and the functions  $\phi$ ,  $\Psi$  and their partial derivatives are evaluated at  $(x_0, y_0)$ . If  $\gamma_2(x_0, y_0)$  is the curvature of the trajectory of the particle which starts from rest at  $(x_0, y_0)$ , then Kasner's theorem states

$$\gamma_2(x_0, y_0) = 1/3\Omega.$$

The result we wish to prove is the following generalization.

**THEOREM:** *If a particle starts from "maximum rest" in an acceleration field of order  $n$ , the initial curvature of the trajectory is  $\frac{n!(n-1)!}{(2n-1)!}$  times the curvature of the line of force through the initial position.*

Again, if  $\phi(x, y)$  and  $\Psi(x, y)$  are the rectangular components of a planar acceleration field of order  $n$  acting at any point  $(x, y)$  and  $t$  is the time, then the equations of motion for a particle of unit mass are

$$\frac{d^nx}{dt^n} = \phi(x, y), \quad \frac{d^ny}{dt^n} = \Psi(x, y). \quad (2)$$

By the phrase "maximum rest," we mean that for  $t = 0$ , the initial conditions are

$$\frac{d^jx}{dt^j} = \frac{d^jy}{dt^j} = 0, \quad 0 < j < n;$$

that is, not merely is the initial velocity zero, but also all the initial accelerations of order up to and including  $n - 2$  are zero.

As before, if the curvature of the lines of force is everywhere zero, then the lines of force are straight lines, and a particle starting from maximum rest at any point must move along the line of force through that point. Thus the trajectory and the line of force coincide and the theorem will be trivially true. Excluding this possibility, we choose a point  $(x_0, y_0)$  at which the curvature of the line of force is not zero. The curvature of the trajectory at  $(x_0, y_0)$  may be calculated from the formula

$$\gamma_n = \frac{\frac{dx}{dt} \frac{d^2y}{dt^2} - \frac{dy}{dt} \frac{d^2x}{dt^2}}{\left[ \left( \frac{dx}{dt} \right)^2 + \left( \frac{dy}{dt} \right)^2 \right]^{3/2}} \quad (3)$$

However, as  $\frac{dx}{dt} = \frac{dy}{dt} = 0$  for  $t = 0$ ,  $\gamma_n$  assumes an indeterminate form and we must have recourse to the theory of limits in order to obtain the desired value. To achieve this, we expand the solution of (2) as power series in the time.

Let the parametric equations of the trajectory be

$$x = a_0 + a_1 t + \dots + \frac{a_r}{r!} t^r + \dots$$

$$y = b_0 + b_1 t + \dots + \frac{b_r}{r!} t^r + \dots$$

where the coefficients  $a_r, b_r$  are to be evaluated. From the initial conditions it follows immediately that

$$a_0 = x_0 \quad a_1 = a_2 = \dots = a_{n-1} = 0, \quad a_n = \varphi(x_0, y_0)$$

$$b_0 = y_0 \quad b_1 = b_2 = \dots = b_{n-1} = 0, \quad b_n = \Psi(x_0, y_0).$$

Differentiate equations (2) with respect to the time. This yields

$$\frac{d^{n+1}x}{dt^{n+1}} = \varphi_x \frac{dx}{dt} + \varphi_y \frac{dy}{dt}, \quad \frac{d^{n+1}y}{dt^{n+1}} = \Psi_x \frac{dx}{dt} + \Psi_y \frac{dy}{dt}.$$

Therefore, for  $t = 0$ ,

$$\frac{d^{n+1}x}{dt^{n+1}} = 0, \quad \frac{d^{n+1}y}{dt^{n+1}} = 0$$

since  $\frac{dx}{dt} = 0$  and  $\frac{dy}{dt} = 0$ . By repeated differentiations we find as initial values for the higher derivatives

$$\frac{d^{n+j}x}{dt^{n+j}} = 0, \quad \frac{d^{n+j}y}{dt^{n+j}} = 0$$

for all values of  $j$  less than  $n$ . However, for  $j = n$ , the appropriate initial values are

$$\frac{d^{2n}x}{dt^{2n}} = \varphi_x\varphi + \varphi_y\Psi, \quad \frac{d^{2n}y}{dt^{2n}} = \Psi_x\varphi + \Psi_y\Psi$$

with  $\varphi, \Psi$  and their partial derivatives evaluated at  $(x_0, y_0)$ . Therefore,

$$\begin{aligned} a_{n+1} = \dots = a_{2n-1} &= 0, & a_{2n} &= \varphi_x\varphi + \varphi_y\Psi \\ b_{n+1} = \dots = b_{2n-1} &= 0, & b_{2n} &= \Psi_x\varphi + \Psi_y\Psi. \end{aligned}$$

and the expansions for the equations of the trajectory are

$$\begin{aligned} x &= x_0 + \frac{\varphi}{n!} t^n + \frac{\varphi_x\varphi + \varphi_y\Psi}{(2n)!} t^{2n} + \text{higher powers} \\ y &= y_0 + \frac{\Psi}{n!} t^n + \frac{\Psi_x\varphi + \Psi_y\Psi}{(2n)!} t^{2n} + \text{higher powers.} \end{aligned}$$

Substituting in formula (3) for the curvature, we obtain

$$\gamma_n = \frac{[\varphi(\Psi_x\varphi + \Psi_y\Psi) - \Psi(\varphi_x\varphi + \varphi_y\Psi)] \left[ \frac{1}{(n-1)!(2n-2)!} - \frac{1}{(n-2)!(2n-1)!} \right] t^{3n-3} + \text{higher powers}}{[\varphi^2 + \Psi^2]^{3/2} \left[ \frac{1}{(n-1)!} \right]^3 t^{3n-3} + \text{higher powers}}.$$

Taking the limit as  $t \rightarrow 0$ , and comparing with formula (1), we have the desired conclusion, namely

$$\gamma_n(x_0, y_0) = \frac{n!(n-1)!}{(2n-1)!} \Omega.$$

This completes the proof.

For the value  $n = 2$ , we obtain as a corollary Kasner's theorem,

$$\gamma_2(x_0, y_0) = 1/3 \Omega.$$

For  $n = 3$ , we find  $\gamma_3 = 1/10 \Omega$ ; for  $n = 4$ ,  $\gamma_4 = 1/35 \Omega$ .

We notice that the ratio of any two successive values of the initial curvature takes the simple form

$$\frac{\gamma_{n+1}}{\gamma_n} = \frac{n+1}{2(2n+1)}.$$

Our general theorem (as in the special<sup>1</sup> case  $n = 2$ ) is found to be valid in ordinary space of any number of dimensions and also in riemannian geometry.

\* Presented to the American Mathematical Society, February, 1942.

<sup>1</sup> See Kasner, E., *Princeton Colloquium Lectures*, 1912, p. 9. Also: *Trans. Am. Math. Soc.*, 1906-1909; *Science*, **75**, 671 (1932); Zurich International Congress of Mathematics, **2**, 180 (1932); *Proc. Nat. Acad. Sci.*, **20**, 130 (1934); Fialkow, A., *Trans. Am. Math. Soc.*, **38**, 89 (1935).

## GENERALIZED TRANSFORMATION THEORY OF ISOTHERMAL AND DUAL FAMILIES\*

BY EDWARD KASNER AND JOHN DE CICCIO

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY AND ILLINOIS INSTITUTE OF  
TECHNOLOGY

Communicated December 31, 1941

*The Isothermal Character Theorem.*—A noteworthy theorem in the theory of functions of a complex variable is that the conformal transformations are the only *point* correspondences which carry every isothermal family of curves into an isothermal family. Kasner's generalization<sup>1</sup> of this result to the lineal-element transformations of the plane is

THEOREM 1. *The group of lineal-element transformations of the real plane which convert every isothermal family into an isothermal family is in cartesian coordinates  $(x, y, \theta)$*

$$X = \phi(x, y), Y = \psi(x, y), \Theta = a\theta + h(x, y), \quad (1)$$

where  $\phi$  and  $\psi$  satisfy the Cauchy-Riemann equations (direct or reverse)

$$\phi_x = \pm\psi_y, \phi_y = \mp\psi_x, \quad (2)$$

and  $h$  is any harmonic function of  $(x, y)$ , and  $a$  is a non-zero constant.

A corollary of this is that the only *contact* lineal-element transformations preserving the isothermal character are the conformal.

This theorem contains most well-known devices for obtaining isothermal families from a given isothermal family. In particular, there appear as special cases the facts that the isogonals and the multiplicatives of an isothermal family each form an isothermal family. However Theorem 1 does not contain the less familiar result that the isoclines of an isothermal family are an isothermal family. Our new generalization to field-element transformations will contain this fact as a special case.

We define a *field-element* of the plane by the number quintuplet  $(x, y, \theta, r, s)$  where  $(x, y)$  are the cartesian coordinates of the point,  $\theta$  is the inclination of the direction,  $r = \theta_x$  and  $s = \theta_y$ . This may be represented geometrically as a lineal-element  $E$  of a field  $F$  together with the tangent flat field of  $F$  at  $E$ .<sup>2</sup>

Any field-element to lineal-element transformation  $T$  may be given by

$$X = \phi(x, y, \theta, r, s), \quad Y = \psi(x, y, \theta, r, s), \quad \Theta = \chi(x, y, \theta, r, s). \quad (3)$$

We shall restrict ourselves to *admissible* transformations  $T$ . Any such transformation  $T$  carries any field with integral curves not all point-unions (stars) into any other such field. For any admissible transformation  $T$ , the functions  $\phi$  and  $\psi$  are *not* functionally dependent. We also exclude the unimportant case where all fields, isothermal or not, are carried into a single isothermal field.

**THEOREM 2.** *The complete set of admissible field-element to lineal-element transformations which carry any isothermal field into an isothermal field is*

$$X = \phi(x, y, r, s), \quad Y = \psi(x, y, r, s), \quad \Theta = a\theta + h(x, y, r, s), \quad (4)$$

where  $\phi$  and  $\psi$  satisfy the Cauchy-Riemann equations (direct or reverse) (for each of the two complex variables  $x + iy$  and  $r - is$ )

$$\phi_x = \pm\psi_y, \quad \phi_y = \mp\psi_x, \quad \phi_r = \mp\psi_s, \quad \phi_s = \pm\psi_r, \quad (5)$$

and  $h$  is any biharmonic function, that is,  $h$  satisfies the four Poincaré partial differential equations of second order

$$h_{xx} + h_{yy} = 0, \quad h_{xr} - h_{ys} = 0, \quad h_{xs} + h_{yr} = 0, \quad h_{rr} + h_{ss} = 0, \quad (6)$$

and  $a$  is a constant.

Next we study the *contact* field-element transformations of the plane. This is the complete set of field-element transformations which carry every field into a field.

**THEOREM 3.** *The entire group of contact field-element transformations which send every isothermal field into an isothermal field consists of the extension of the Kasner group (1), and the set of transformations*

$$\begin{aligned} r &= \phi(X, Y, x, y), \quad s = \psi(X, Y, x, y), \\ \Theta &= a[\theta - \int \{r dx + s dy\}] + h(X, Y), \\ R &= -a \int \{r_X dx + s_X dy\} + h_X, \quad S = -a \int \{r_Y dx + s_Y dy\} + h_Y \end{aligned} \quad (7)$$

where  $\phi$  and  $\psi$  satisfy the Cauchy-Riemann equations

$$\phi_x = \psi_y, \quad \phi_y = -\psi_x, \quad \phi_X = \mp\psi_Y, \quad \phi_Y = \pm\psi_X, \quad (8)$$

and  $h$  is a harmonic function of  $(X, Y)$ , and  $a$  is a non-zero constant.

The field-element to lineal-element transformation by which every field is carried into its isocline field is

$$X = x, Y = y, \Theta = -\arctan r/s. \quad (9)$$

This transformation obviously comes under Theorem 2 but *not* under Theorem 3 since it is not a contact transformation.

*The Dual-Isothermal Character Theorem.*—Any system of  $\infty^1$  curves which under the group of equilong transformations of the plane can be carried into the  $\infty^1$  point-unions (stars) of a straight line has been called *dual-isothermal* by Kasner. A field, given in hessian coordinates  $(u, v, w)$  by  $w = w(u, v)$ , is dual-isothermal if and only if  $w$  satisfies the dual-Laplace equation  $w_{vv} = 0$ , that is, if  $w = v\alpha(u) + \beta(u)$ .

By the definition, any equilong transformation carries every dual-isothermal family into a dual-isothermal family. The generalization<sup>5</sup> of this result to lineal-element transformations of the plane is

**THEOREM 4.** *The group of lineal-element transformations which carry every dual-isothermal family into a dual-isothermal family is*

$$U = \phi, V = \frac{a_2v + b_2w + c_2}{a_1v + b_1w + c_1}, W = \frac{a_3v + b_3w + c_3}{a_1v + b_1w + c_1}, \quad (10)$$

where  $\phi, a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3, c_3$  are functions of  $u$  only.

A corollary of this is that the *contact* lineal-element transformations preserving the isothermal character are

$$U = \phi(u), V = v\psi(u) + \chi(u), W = \frac{w\psi + v\psi_u + \chi_u}{\phi_u}. \quad (11)$$

This is a group of line transformations larger than the equilong group. We obtain the equilong group by imposing the condition  $\psi = \phi_u$ .

A field-element of the plane may be defined by the number quintuplet  $(u, v, w, p, q)$ , where  $(u, v, w)$  are the hessian coordinates of the lineal element  $E$  and  $p = w_u$  and  $q = w_v$ .

Any field-element to lineal-element transformation  $T$  may be given by

$$U = \phi(u, v, w, p, q), V = \psi(u, v, w, p, q), W = \chi(u, v, w, p, q). \quad (12)$$

We restrict ourselves to admissible transformations  $T$ . Any such transformation  $T$  carries any field with integral curves not all straight lines into any other such field. For any admissible transformation  $T$ , the functions  $\phi$  and  $\chi$  are *not* functionally dependent. The unimportant case where all fields, dual-isothermal or not, are carried into a single dual-isothermal field is also excluded.

**THEOREM 5.** *The complete set of admissible field-element to lineal-element transformations which send any dual-isothermal field into a dual-isothermal field is*

$$U = \phi, V = \frac{a_2 v + b_2 p + c_2}{a_1 v + b_1 p + c_1}, W = \frac{a_3 v + b_3 p + c_3}{a_1 v + b_1 p + c_1}, \quad (13)$$

where  $\phi, a_1, b_1, c_1, a_2, b_2, c_2, a_3, b_3, c_3$  are functions of  $u, q$ , and  $(w - qv)$  only.

Next we study the contact field-element transformations of the plane.

**THEOREM 6.** *The entire group of contact field-element transformations which convert every dual-isothermal field into a dual-isothermal field is*

$$\begin{aligned} U &= U(u, q, \omega), Q = Q(u, q, \omega), W = QV + \Omega(u, q, \omega), \\ V &= \frac{v(U_u \Omega_\omega - U_\omega \Omega_u) + p(U_\omega \Omega_q - U_q \Omega_\omega) + (U_q \Omega_u - U_u \Omega_q)}{v(U_\omega Q_u - U_u Q_\omega) + p(U_q Q_\omega - U_\omega Q_q) + (U_u Q_q - U_q Q_u)}, \\ P &= \frac{v(\Omega_\omega Q_u - \Omega_u Q_\omega) + p(\Omega_q Q_\omega - \Omega_\omega Q_q) + (\Omega_u Q_q - \Omega_q Q_u)}{v(U_\omega Q_u - U_u Q_\omega) + p(U_q Q_\omega - U_\omega Q_q) + (U_u Q_q - U_q Q_u)} \end{aligned} \quad (14)$$

where  $\omega = w - qv$ .

The field-element transformation whereby every field is carried into its dual-isocline field is

$$U = u, V = v, W = -p/q. \quad (15)$$

This comes under Theorem 5 but not under Theorem 6 since it is not a contract transformation.<sup>4</sup>

\* Presented to the American Mathematical Society, February, 1942.

<sup>1</sup> Kasner, "Lineal Element Transformations Which Preserve the Isothermal Character," *Proc. Nat. Acad. Sci.*, 27, 406-409 (1941). Conformal mapping obviously converts isothermal families into isothermals. But the converse, that no other point transformations can be successful, does not seem to appear in the standard literature; and was first obtained as a corollary of the search for element transformations in that paper.

<sup>2</sup> Kasner and De Cicco, "The Geometry of Turbines, Flat Fields, and Differential Equations," *Am. Jour. Math.*, 59, 545-563 (1937).

<sup>3</sup> De Cicco, "Lineal Element Transformations Which Preserve the Dual-Isothermal Character," *Proc. Nat. Acad. Sci.*, 27, 409-412 (1941).

<sup>4</sup> Whereas the isocline field of any field  $F$  is determined by the point-unions on the  $\infty^1$  equiparallel series of  $F$ , the dual-isocline field of any field  $F$  is defined with respect to a given linear turbine  $T$  as follows: Let  $S_0$  be the series of intersection of  $F$  and any non-linear flat field whose central element  $G$  is on  $T$ . The line-unions of these  $\infty^1$  coflat series  $S_0$  form the dual-isocline field of  $F$ . Thus if  $(u, v, w)$  represent the *equilong* coordinates of a lineal-element, then the correspondence (14) represents the dual-isocline field of any field with respect to the linear turbine  $T$  whose elements all lie on the negative  $y$ -axis.



## CONCERNING SEPARABILITY

BY R. L. MOORE

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF TEXAS

Communicated December 23, 1941

Axiom 1 of the author's Foundations of Point Set Theory<sup>1</sup> is as follows.

AXIOM 1. *There exists a sequence  $G_1, G_2, G_3, \dots$  such that (1) for each  $n$ ,  $G_n$  is a collection of regions covering  $S$ , (2) for each  $n$ ,  $G_{n+1}$  is a subcollection of  $G_n$ , (3) if  $R$  is any region whatsoever,  $X$  is a point of  $R$  and  $Y$  is a point of  $R$  either identical with  $X$  or not, then there exists a natural number  $m$  such that if  $g$  is any region belonging to the collection  $G_m$  and containing  $X$  then  $\bar{g}$  is a subset of  $(R - Y) + X$ , (4) if  $M_1, M_2, M_3, \dots$  is a sequence of closed point sets such that, for each  $n$ ,  $M_n$  contains  $M_{n+1}$  and, for each  $n$ , there exists a region  $g_n$  of the collection  $G_n$  such that  $M_n$  is a subset of  $\bar{g}_n$  then there is at least one point common to all the point sets of the sequence  $M_1, M_2, M_3, \dots$ .*

Axiom 0 states that every region is a point set.

In this paper the following theorem will be established.

THEOREM 1. *If Axioms 0 and 1 hold true and there do not exist uncountably many mutually exclusive domains then space is separable.*

*Proof.* There exists a well-ordered sequence  $\alpha$  whose terms are the regions of  $G_1$ . There exists a well-ordered subsequence  $\alpha_1$  of  $\alpha$  such that (1) the first term of  $\alpha_1$  is the first term of  $\alpha$ , (2) if  $\beta$  is a well-ordered subsequence of  $\alpha_1$  distinct from  $\alpha_1$  and such that every term of  $\alpha_1$  which precedes a term of  $\beta$  belongs to  $\beta$  then (a) there exists a region  $R$  belonging to  $\alpha$  and intersecting no region of  $\beta$  and (b) the first such region  $R$  in  $\alpha$  is the first term of  $\alpha_1$  following all the terms of  $\beta$ . Let  $H_1$  denote the collection of all regions of the sequence  $\alpha_1$ . The regions of  $H_1$  are mutually exclusive and every point either belongs to  $H_1^*$  or is a limit point of it.<sup>2</sup> Similarly there exists a collection  $H_2$  of mutually exclusive regions such that (1)  $H_2$  is a subcollection of  $G_2$ , (2) if  $x$  is a region of  $H_2$ ,  $\bar{x}$  is a subset of some region of  $H_1$ , (3) if  $x$  is a region of  $H_1$  every point of  $x$  belongs to, or is a limit point of, the sum of all the regions of  $H_2$  which are subsets of  $x$ . This process may be continued. Thus there exists a sequence  $H_1, H_2, H_3, \dots$  such that, for each  $n$ , (1)  $H_n$  is a subcollection of  $G_n$ , (2) if  $x$  is a region of  $H_{n+1}$ ,  $\bar{x}$  is a subset of some region of  $H_n$ , (3) the regions of  $H_n$  are mutually exclusive, (4) if  $x$  is a region of  $H_n$  every region that intersects  $x$  intersects the sum of all the regions of  $H_{n+1}$  which are subsets of  $x$ .

Let  $H$  denote the set of all regions  $h$  such that, for some  $n$ ,  $h$  belongs to  $H_n$ . Since, for each  $n$ , the regions of  $H_n$  are mutually exclusive, the collection  $H$  is countable. Suppose  $R$  is a region. There exists a region  $R_1$  such that  $\bar{R}_1$  is a subset of  $R$ . The region  $R_1$  intersects some region  $x_1$  of  $H_1$ . It also intersects some region  $x_2$  of  $H_2$  which is a subset of  $x_1$ . This process may be continued. Thus there exists a sequence of regions  $x_1, x_2, x_3, \dots$  such

that, for each  $n$ ,  $x_n$  belongs to  $H_n$ , intersects  $R_1$  and contains  $\bar{x}_n + 1$ . But, for each  $n$ ,  $x_n$  belongs to  $G_n$ . Hence the regions of this sequence have a point  $O$  in common. Suppose  $O$  does not belong to  $\bar{R}_1$ . There exists a number  $n$  such that if  $g$  is any region of  $G_n$  containing  $O$  then  $\bar{g}$  is a subset of the domain  $S - \bar{R}_1$ , contrary to the fact that  $x_n$  contains  $O$ , belongs to  $G_n$  and intersects  $\bar{R}_1$ . Hence  $O$  belongs to  $\bar{R}_1$  and therefore to  $R$ . It follows that, for some  $n$ ,  $x_n$  is a subset of  $R$ . Thus every region contains some region of  $H$ . Since  $H$  is countable it follows that space is separable.

Let  $Q$  denote the axiom obtained from Axiom 1 by replacing stipulation (4) concerning  $G_1, G_2, G_3, \dots$  in the statement of that Axiom by the stipulation that if  $x$  and  $y$  are regions such that  $\bar{x}$  is a subset of  $y$  and  $g_1, g_2, g_3, \dots$  is a sequence of regions such that, for each  $n$ ,  $g_n$  contains  $\bar{g}_n + 1$ , belongs to  $G_n$  and intersects  $x$  then, for some  $n$ ,  $g_n$  is a subset of  $y$ . By an easily made modification of the above proof, it may be shown that Theorem 1 remains true if, in the statement of its hypothesis, "1" is replaced by " $Q$ ."

Clearly Axiom  $Q$  holds true in every metric space. But, in the presence of Axiom  $O$ , it is weaker than Axiom 1. It has been shown, by John H. Roberts,<sup>3</sup> that Axiom 1 holds true for no metric spaces which are not complete. But there exist both separable and non-separable spaces which satisfy Axioms 0, 1, 2, 3, 4, 5<sub>1</sub>, 5<sub>2</sub> and 6 and which are not metric. Consider the following examples.

*Example 1.* In a three-dimensional Euclidean space  $E$ , let  $\beta$  denote a definite straight line and let  $O$  denote a definite point lying on it. Let  $G$  denote the set of all point sets ("half-planes")  $g$  such that, for some plane  $M$  containing  $\beta$ ,  $g$  is the sum of  $\beta$  and a component of  $M - \beta$ . There is a one to one correspondence between the points of  $\beta$  and the half-planes of the set  $G$ . For each point  $P$  of  $\beta$  let  $T(P)$  denote the half-plane of  $G$  with which it is paired in this correspondence. For each point  $P$  of  $\beta$  let  $W_P$  denote the set of all straight line rays which start from  $P$  and lie, except for  $P$ , wholly in  $T(O) - \beta$ . If  $P$  is a point of  $\beta$  and  $a$  is a ray of  $W(P)$ , let  $Z(a)$  denote the image of  $a$  under a rotation around  $\beta$  that throws the plane  $T(O)$  into the plane  $T(P)$ . If  $P$  is a point of  $\beta$ ,  $a$  and  $b$  are rays of  $W_P$  and  $\epsilon$  is a positive number, let  $R_{Pab\epsilon}$  denote a set such that  $x$  is an element of it only if either (1)  $x$  is the open curve  $y + Z(y)$  for some ray  $y$  of the set  $W_P$  such that  $y - P$  is separated from  $\beta - P$  by  $a + b$  in the half-plane  $T(O)$  or (2)  $x$  is a point which is at a distance less than  $\epsilon$  from  $P$  and which is separated from  $\beta - P$  by  $a + b$  in  $T(O)$  or from  $\beta - P$  by  $Z(a) + Z(b)$  in  $T(P)$ .

Let  $\Sigma$  denote a space such that (1)  $X$  is a point of  $\Sigma$  if and only if either  $X$  is a point of  $E - \beta$  or, for some point  $P$  of  $\beta$ ,  $X$  is the open curve  $a + Z(a)$  for some ray  $a$  of  $W_P$ , (2)  $R$  is a region of  $\Sigma$  if and only if either, in some half-plane of the set  $G$ ,  $R$  is the interior of some circle containing no point of  $\beta$  or  $R$  is  $R_{Pab\epsilon}$  for some  $P$ ,  $a$ ,  $b$  and  $\epsilon$ . This space satisfies all of the above mentioned axioms. But it is not separable.

*Example 2.* In a Euclidean plane  $E$ , let  $\beta$  denote a definite straight line, let  $A$  and  $B$  denote two definite points of  $\beta$  and let  $N$  denote a "half-plane" which is the sum of  $\beta$  and a definite one of the two components of  $E - \beta$ . If  $P$  is a point of  $\beta$  let  $A_P$  and  $B_P$  denote points of  $\beta$  such that the straight line rays  $PA_P$  and  $PB_P$  have only  $P$  in common and the common part of the straight line rays  $PB_P$  and  $AB$  is a ray. Let  $C_P$  denote a point of  $N$  such that the straight line ray  $PC_P$  is perpendicular to  $\beta$ . Let  $W_P$  denote the set of all straight line rays which start from  $P$  and lie, except for  $P$ , in the interior of the right angle  $C_PPB_P$ . For each ray  $a$  of  $W_P$ , let  $Z(a)$  denote a straight line ray lying, except for  $P$ , in the interior of the right angle  $C_PPA_P$  and such that the acute angle whose sides are  $Z(a)$  and  $PC_P$  is congruent to the one whose sides are  $a$  and  $PC_P$ .

If  $P$  is a point of  $\beta$ ,  $a$  and  $b$  are rays of  $W_P$  and  $\epsilon$  is a positive number, let  $R_{Pab\epsilon}$  denote a set such that  $x$  is an element of it only if either (1)  $x$  is the open curve  $y + Z(y)$  for some ray  $y$ , of the set  $W_P$ , such that  $y - P$  lies in the interior of the acute angle whose sides are  $a$  and  $b$  or (2)  $x$  is a point which is at a distance less than  $\epsilon$  from  $P$  and which lies either in the interior of the acute angle whose sides are  $a$  and  $b$  or in the interior of the one whose sides are  $Z(a)$  and  $Z(b)$ .

If  $P$  is a point of  $\beta$ ,  $a$  is a ray of the set  $W_P$  and  $\epsilon$  is a positive number, let  $R_{Pa\epsilon}$  denote a set such that  $x$  is an element of it only if either (1)  $x$  is the ray  $PC_P$  or (2)  $x$  is the open curve  $y + Z(y)$  for some ray  $y$  of the set  $W_P$  such that  $y - P$  lies in the interior of the acute angle whose sides are  $a$  and  $PC_P$  or (3)  $x$  is a point which is at a distance less than  $\epsilon$  from  $P$  and which, in the half-plane  $N$ , is separated from  $\beta - P$  by the open curve  $a + Z(a)$ .

Let  $\Sigma$  denote a space such that (1)  $X$  is a point of  $\Sigma$  if and only if either  $X$  is a point of  $E - \beta$  or, for some point  $P$  of  $\beta$ ,  $X$  is  $PC_P$  or the open curve  $a + Z(a)$  for some ray  $a$  of  $W_P$ , (2)  $R$  is a region of  $\Sigma$  if and only if either, in the half-plane  $N$ ,  $R$  is the interior of some circle containing no point of  $\beta$  or  $R$  is  $R_{Pab\epsilon}$  for some  $P$ ,  $a$ ,  $b$  and  $\epsilon$  or  $R$  is  $R_{Pa\epsilon}$  for some  $P$ ,  $a$  and  $\epsilon$ . The space  $\Sigma$  satisfies Axioms 0-4, 5<sub>1</sub>, 5<sub>2</sub> and 6. Furthermore it is separable. But it is not completely separable.

<sup>1</sup> *Am. Math. Soc. Colloquium Pub.*, Vol. XIII, New York (1932). The letter  $S$  denotes the set of all points.

<sup>2</sup> The symbol  $H^*$  denotes the sum of all the point sets of the collection  $H$ .

<sup>3</sup> Roberts, J. H., *Bull. Am. Math. Soc.*, 38, 835-838 (1932).

THE COEFFICIENTS OF STIRLING'S SERIES FOR  $\log \Gamma(x)$ 

BY HORACE S. UHLER

SLOANE PHYSICS LABORATORY, YALE UNIVERSITY

Communicated December 29, 1941

The asymptotic series

$$\log \Gamma(x) = (\log 2\pi)/2 + (x - 1/2) \log x - x + \sum_{m=1}^{\infty} (c_m/x^{2m-1}) + R,$$

where  $c_m \equiv (-1)^{m-1} B_m / [(2m-1)(2m)]$ , obviously requires for its evaluation adequate approximations to the numerical values of  $\log 2\pi$ ,  $\log x$  and  $c_m$ . The desire to apply this series to the extension of my table<sup>1</sup> of  $1/n!$  ( $\Gamma(n+1) = n!$ ,  $n$  integral), for higher values of  $n$  led to the original calculation<sup>2</sup> of  $\log \pi$  to 214 significant figures, to the computation of two 137-place tables of natural logarithms for factors of the form  $1 \pm a \cdot 10^{-b}$ , and finally to the calculation of the table of coefficients ( $c_m$ ) presented below. Incidentally the table given by Duarte<sup>3</sup> is insufficient for my purposes because it was only designed to furnish about 42 significant figures. This table also has the disadvantage of greatly increasing the number of figures to be written in solving many problems because it is founded upon the base 10 instead of  $e$ .

TABLE 1

$c_1 = +0.083$	$c_2 = -0.0027$
$c_3 = +0.00079 \ 3650$	$c_4 = -0.00059 \ 52380$
$c_5 = +0.00084 \ 1750$	$c_6 = -0.00191 \ 7526$
$c_7 = +0.00641 \ 025$	$c_8 = -0.02955 \ 06535 \ 94771 \ 24183 \ 00$
$c_9 = +0.17964 \ 43723 \ 68830 \ 57316 \ 49384 \ 90015 \ 88939 \ 66943 \ 50254 \ 72177$	
	$17496 \ 35526 \ 72531 \ 00070 \ 43753 \ 17378 \ 41335 \ 36455 \ 51787 \ 96664$
	$86477 \ 63198 \ 84679 \ 01780 \ 6n$
$c_{10} = -1.39243 \ 22169 \ 05901 \ 11642 \ 7$	
$c_{11} = +13.40286 \ 40441 \ 68391 \ 99447 \ 89510 \ 00690 \ 13112 \ 49137 \ 33609$	
	$38578 \ 32988 \ 26777 \ 08764 \ 665$
$c_{12} = -156.84828 \ 46260 \ 02017 \ 30636 \ 51324 \ 52088 \ 97382 \ 81042 \ 62886$	
	$87158 \ 25237 \ 56436 \ 79991 \ 50607$
$c_{13} = +2193.103$	
$c_{14} = -36108.77125 \ 87249 \ 89357 \ 17326 \ 52192 \ 42230 \ 73648 \ 36100 \ 46828$	
	$43763 \ 30353 \ 34184 \ 75947 \ 21157 \ 93954 \ 87441 \ 46445 \ 29587 \ 05832$
	$26905 \ 1N$
$c_{15} = +6 \ 91472.26885 \ 13180 \ 67108 \ 39525 \ 07756 \ 73467 \ 55333 \ 40716$	
	$87798 \ 05042 \ 31894 \ 66571 \ 00160 \ 99337 \ 56763 \ 56766 \ 45687 \ 76915$
	$82919 \ 61P$
$c_{16} = -152 \ 38221.53940 \ 74161 \ 92283 \ 36495 \ 88867 \ 80518 \ 65907 \ 65338$	
	$39342 \ 18848 \ 82985 \ 45224 \ 54142 \ 94750 \ 15812 \ 77672 \ 35926 \ 62871$
	$600P$

(Continued on page 60)

TABLE 1. (Continued)

[illegible]

TABLE 1 (Continued)

$c_{49}$	= +	11906 21023 08902 26457 68481 61768 7138	$p$ (39)
$c_{50}$	= -	28668 93896 02966 73696 22642 05419 01	$n$ (43)
$c_{51}$	= +	71893 07802 33729 86439 28443 71393	$p$ (47)
$c_{52}$	= -	18760 69343 05046 71847 89546 2727	$N$ (51)
$c_{53}$	= +	50904 91469 07510 70481 32915 19	$N$ (55)
$c_{54}$	= -	14351 42882 84321 71093 65153 5	$N$ (59)
$c_{55}$	= +	42009 05750 66658 69407 2472	$p$ (63)
$c_{56}$	= -	12758 58127 22475 75908 433	$N$ (67)
$c_{57}$	= +	40177 56840 36218 18804 6	$n$ (71)
$c_{58}$	= -	13110 13631 20068 34587	$P$ (75)
$c_{59}$	= +	44299 95653 69782 596	$P$ (79)
$c_{60}$	= -	15492 14069 51735 54	$n$ (83)
$c_{61}$	= +	56037 64855 62735	$p$ (87)
$c_{62}$	= -	20953 92414 8581	(91)
$c_{63}$	= +	80952 90300 56	$P$ (95)
$c_{64}$	= -	32296 33556 0	$P$ (99)
$c_{65}$	= +	13298 63917	$P$ (103)
$c_{66}$	= -	56491 165	$N$ (107)
$c_{67}$	= +	24743 78	$n$ (111)
$c_{68}$	= -	11170 2	$P$ (115)
$c_{69}$	= +	5195	$n$ (119)
$c_{70}$	= -	249	$n$ (123)
$c_{71}$	= +	12	$P$ (127)

$$\log_e (100!) =$$

363.73937 55555 63490 14407 99933 69655 63802 78239 21062  
 88727 47276 79448 87677 59444 47979 01991 41010 00241 97254  
 93196 15773 55972 29305 31198 01503 48915 04259 44052 15183  
 63651 214

$$100! =$$

$2^{97} \cdot 3^{48} \cdot 5^{24} \cdot 7^{16} \cdot 11^9 \cdot 13^7 \cdot 17^5 \cdot 19^4 \cdot 23^4 \cdot 29^3 \cdot 31^3 \cdot 37^2 \cdot 41^2 \cdot 43^2 \cdot 47^2 \cdot 53 \cdot 59 \cdot 61 \cdot 67 \cdot 71 \cdot 73 \cdot 79 \cdot$   
 $83 \cdot 89 \cdot 97 =$

933 26215 44394 41526 81699 23885 62667 00490 71596 82643  
 81621 46859 29638 95217 59999 32299 15608 94146 39761 56518  
 28625 36979 20827 22375 82511 85210 91686  $4 \times 10^{24}$

All of the values of  $c_m$  here tabulated were worked out very carefully and independently by an assistant and by me. We used two different machines located in widely separated buildings. As far as  $m = 62$  the assistant transcribed the values of the Bernoulli numbers ( $B_m$ ) as calculated and expressed as repeating decimals<sup>4</sup> by Adams while I used the improper fractions<sup>5</sup> having the same author. For values of  $m$  greater than 62 the numerators and denominators in Serebrennikoff's earlier table<sup>6</sup> were employed by both of us. The reliability of the borrowed numbers as printed is validated by the following quotation from D. H. Lehmer:<sup>7</sup> "This check assures the correctness not only of the present table but also of the tables of Adams and Serebrennikoff inasmuch as their values were used in computing the table of  $G_n$ ." The table of  $c_m$  as a whole was checked by substituting  $x = n = 100$  and finding  $\neq \log (100!)$ . The antilogarithms were evaluated by means of my tables of  $\log (1 \pm a \cdot 10^{-b})$ , and they agreed to 141 figures with the previously established values of  $100!$  and  $1/100!$ . These constants

may be found, respectively, at the end of the present new table and under the entry  $n = 100$  in the table of reciprocals of factorials.<sup>1</sup>

With one minor exception all of the values of  $c_m$  whose recurring periods do not end within the range of the table contain the right number of figures to cause the corresponding terms in Stirling's series to end at the 155th decimal place when  $n = 100$ .  $c_{37}$  has been extended by one digit in order to exhibit both ends of the first period of its repeating decimal. Beginning with  $c_{40}$  the number of omitted figures between the last recorded digit and the decimal point is indicated in each case by the number which is enclosed between parentheses.

Estimates of the errors incurred by rounding off the terminal digits may be formed from the somewhat refined convention here introduced. Let  $c'_m$  represent the unlimited positive or negative number which, when added to the finitely terminated tabular value of  $|c_m|$ , will give the true arithmetical value of  $B_m/[(2m-1)(2m)]$ , that is, the complete value of infinite extent.  $c'_m$  is to be expressed as a decimal such that unit's place coincides with the final digit  $f_m (= 0, 1, 2, \dots, 9)$  of  $c_m$  as recorded. When  $c'_m$  falls within the narrow range from  $-0.02$  to  $+0.02$  no letter follows  $f_m$  in the table. If  $c'_m$  lies between  $-0.50$  and  $-0.25$  the symbol  $N$  succeeds the digit  $f_m$ . If  $c'_m$  is greater than  $-0.25$  and less than  $-0.02$  the letter  $n$  is appended. According to the same scheme the indices  $p$  and  $P$  correspond, respectively, to the ranges from  $+0.02$  to  $+0.25$  and from  $+0.25$  to  $+0.50$ . It should now be clear that the symbols are intended to suggest negative and positive corrections, that the capital letters indicate the quarter intervals which are most remote from the center of reference  $f_m$ , and that  $f_m$  is an enhanced digit whenever it is followed by  $N$  or  $n$ .

<sup>1</sup> Uhler, H. S., *Trans. Conn. Acad. Arts. Sci.*, 32, 381-434 (1937).

<sup>2</sup> Uhler, H. S., *Proc. Nat. Acad. Sci.*, 24, 23-30 (1938).

<sup>3</sup> Duarte, F.-J., *Nouvelles Tables de Log n!*, Geneva, XIII (1927).

<sup>4</sup> *Scientific Papers of John Couch Adams*, 1, 455-458 (1896).

<sup>5</sup> *Ibid.*, 453.

<sup>6</sup> Serebrennikoff, S. Z., *Mém. Acad. Imp. Sci. St.-Petersbourg*, 16, No. 10 (1905).

<sup>7</sup> Lehmer, D. H., *Duke Math. Jour.*, 2, 460-464 (1936).

## OSCILLATIONS OF THE DERIVATIVES OF A FUNCTION

BY A. C. SCHAEFFER

DEPARTMENT OF MATHEMATICS, STANFORD UNIVERSITY

Communicated January 2, 1942

It is known that if a function is infinitely often differentiable in an open interval and all the derivatives are positive then the function is analytic in



the interval.<sup>1</sup> As a possible generalization of this result Pólya raised the following question:<sup>2</sup> If in an open interval a function is infinitely often differentiable and no derivative changes sign more than a fixed bounded number of times, is the function analytic in the interval? The present paper contains an answer to this question and to several related questions. These results are consequences of a fundamental inequality which relates the magnitude of the first derivative of a function to the number of times some higher derivative changes sign.

**THEOREM I.** *In an interval  $a - L < x < a + L$  let  $f(x) \in C^n$ ,  $n \geq 2$ , and let  $|f(x)| \leq M$ . If*

$$|f'(a)| \geq (20n)^{2n} M/L \quad (1)$$

*then  $f^{(n)}(x)$  changes sign at least  $n - 1$  times in the interval.*

The idea in the proof of this theorem is that if the first derivative of a function is "large" at some point then there will be two near-by points where the second derivative is "large" and with different signs. There will then be three near-by points where the third derivative is "large" and with alternating sign, and so on.

**THEOREM II.** *In the interval  $a < x < b$  let  $f(x) \in C^\infty$  and let no derivative change sign more than a fixed bounded number of times. Then  $f(x)$  is analytic in the interval.*

Consider the case in which the interval is  $-1 < x < 1$  and  $|f(x)| \leq 1$ . Let no derivative change sign more than  $k - 2$  times in the interval. It is shown by induction that for  $n = 0, 1, 2, \dots$  and for  $-1 < x < 1$ ,

$$|f^{(n)}(x)| \leq \alpha^n n! (1 - |x|)^{-n} \quad (2)$$

where

$$\alpha = 3(20k)^{2k}.$$

If (2) is true for some  $n$ , and  $x_0$  is some point in  $(-1, 1)$ , then in the subinterval  $|x - x_0| \leq (1 - |x_0|)(n + 1)^{-1}$  (2) implies that

$$|f^{(n)}(x)| \leq 3\alpha^n n! (1 - |x_0|)^{-n};$$

and  $f^{(n+k)}(x)$  does not change sign more than  $k - 2$  times in this subinterval. The induction is then easily completed by use of Theorem I. It follows from (2) that the Taylor's expansion of  $f(x)$  about any point in the interval  $-1 < x < 1$  converges to  $f(x)$  in a complete neighborhood of the point.

A question which naturally arises is whether it is possible to let the number of sign changes of  $f^{(n)}(x)$  tend to infinity with  $n$ , but sufficiently slowly, and still conclude that the function is analytic. While Theorem I leaves the question open for a finite interval it does say something in this direction for an infinite interval.

**THEOREM III.** *Let  $f(x) \in C^\infty$  in  $-\infty < x < \infty$ . Suppose there are finite*



constants  $\gamma > 1$  and  $L > 0$  such that in every interval of length  $L$  the number of sign changes of  $f^{(n)}(x)$  is less than  $(\log n) (\log \log n)^{-\gamma}$  for  $n > n_0$ . If

$$\overline{\lim}_{x \rightarrow \pm \infty} \frac{1}{|x|} \log |f(x)| < \infty$$

then  $f(x)$  is an entire function of order at most one.

In the case in which the number of sign changes of  $f^{(n)}(x)$  is uniformly bounded in every interval of length, say,  $L$ , it is possible to obtain more precise information concerning the function.

**THEOREM IV.** Let  $f(x) \in C^\infty$  in  $-\infty < x < \infty$ . Suppose there are constants  $a = [a] > 0$  and  $L > 0$  such that  $f^{(n)}(x)$  does not change sign more than  $a$  times in any interval of length  $L$ ,  $n = 1, 2, 3, \dots$ . If

$$\overline{\lim}_{x \rightarrow \pm \infty} \frac{1}{|x|} \log |f(x)| < \infty$$

then  $f(x)$  is an entire function of exponential type.

An entire function is said to be of exponential type if for some constant  $\rho$  it satisfies

$$f(z) = O(e^{\rho|z|})$$

uniformly in every direction as  $|z| \rightarrow \infty$ . In Theorem IV, if the function is bounded on the real axis, the proper bound for the constant  $\rho$  can be determined in terms of  $a$  and  $L$ .

Pólya and Wiener have shown<sup>1</sup> that if a function belongs to  $C^\infty$  over  $-\infty < x < \infty$ , is periodic with period  $p$ , and no derivative changes sign more than  $q = [q]$  times in any interval of length  $p$ , then the function is a trigonometric polynomial. It is clear that this result is a consequence of Theorem IV. In a yet unpublished paper Tamarkin has obtained closely related results for functions which are not periodic, but are of integrable square on the real axis.

<sup>1</sup> Bernstein, S., *Leçons sur les propriétés extrémales . . .*, Paris, 1926, p. 190.

<sup>2</sup> At the Stanford University Symposium, August 12, 1941.

<sup>3</sup> The functions considered are real for real values of the variable.

<sup>4</sup> Pólya, G., and Wiener, N., in a paper to appear in the *Trans. Amer. Math. Soc.*

PROCEEDINGS  
OF THE  
NATIONAL ACADEMY OF SCIENCES

Volume 28

March 15, 1942

Number 3

Copyright 1942 by the National Academy of Sciences

*HYPOXANTHINE, A GROWTH SUBSTANCE FOR PHYCOMYCES*

BY WILLIAM J. ROBBINS AND FREDERICK KAVANAGH

NEW YORK BOTANICAL GARDEN AND DEPARTMENT OF BOTANY, COLUMBIA UNIVERSITY

Communicated January 23, 1942

In previous papers from this laboratory an unidentified growth substance for *Phycomyces* widely distributed in products of natural origin was described as factor  $Z_1$ .<sup>1</sup> Guanine was found to have the same effect as factor  $Z_1$ , but was not considered to be identical with it because the activity of the former was destroyed by treatment with nitrous acid while that of the latter was unaffected. Xanthine, as might be anticipated from the effect of nitrous acid on the activity of guanine, was found to be inactive. Choline, adenine, thymine, uracil and cytosine also were ineffective.<sup>2</sup> Further experiments have been carried out in an attempt to identify factor  $Z_1$ .

The procedure followed was that described in the previous paper.<sup>2</sup> The growth of *Phycomyces* during a period of 72 hours at 26°C. was observed in 25 ml. of a basal mineral-dextrose solution containing asparagine and thiamine to which the various compounds or preparations tested for factor  $Z_1$  activity were added alone and in combination with a  $D_R$  fraction<sup>3</sup> prepared from white potato tubers. We have used also the 48-hour development in 10 ml. of solution, a modification which economizes time and material.

The following were tested for factor  $Z_1$  activity: monomethyl guanosine, monoacetone guanosine, guanosine picrate, guanylic acid, sodium guanylate, isoguanine sulfate, sodium inosine, hypoxanthine, vicine, divicine sulfate, barium inosinate and 2-amino uric acid (all obtained through the courtesy of the Rockefeller Institute for Medical Research), hypoxanthine (prepared in our laboratory by treating adenine with nitrous acid), a concentrate of folic acid supplied by R. J. Williams and one obtained from W. H. Peterson, uric acid, amygdalin, allantoin, theobromine, theophyllin, theocin,<sup>4</sup> adenine, guanine and xanthine.

Guanine and both samples of hypoxanthine were active and their activity on an equimolecular basis was approximately the same (table 1). None of the other compounds or the two concentrates of folic acid was

effective, at least in amounts of the same order of magnitude as the effective quantities of guanine or hypoxanthine. Hypoxanthine<sup>6</sup> was isolated in crystalline form as the silver salt (Fig. 1) from a Ca fraction from white potato tubers which had been treated with nitrous acid and electrodialed and which showed factor  $Z_1$  activity. From 60 ml. of the solution 15.5 mg. of the silver salt of hypoxanthine were obtained. The substance isolated from this salt was active. On this basis it was estimated that the factor  $Z_1$  activity of the extract from which the hypoxanthine was isolated could be accounted for by its hypoxanthine content. It would appear, therefore, that the active substance present in the potato extract after treatment with nitrous acid was hypoxanthine.

TABLE 1

THE EFFECT OF GUANINE AND HYPOXANTHINE IN THE PRESENCE OF THIAMINE AND A  $D_R$  FRACTION FROM POTATO TUBERS. NOTICE THAT BOTH COMPOUNDS ARE ABOUT EQUALLY EFFECTIVE IN INCREASING THE BENEFICIAL ACTION OF THE  $D_R$  FRACTION

ADDITIONS TO 10 ML. OF BASAL SOLUTION PLUS THIAMINE	DRY WT. 2 CULTURES MG.	ADDITIONS TO 10 ML. OF BASAL SOLUTION PLUS THIAMINE	DRY WT. 2 CULTURES MG.
10 $\mu$ g. guanine HCl plus $D_R$ fraction	63	6.6 $\mu$ g. hypoxanthine plus $D_R$ fraction	64
1 $\mu$ g. guanine HCl plus $D_R$ fraction	61	0.66 $\mu$ g. hypoxanthine plus $D_R$ fraction	67
0.1 $\mu$ g. guanine HCl plus $D_R$ fraction	43	0.066 $\mu$ g. hypoxanthine plus $D_R$ fraction	42
None	0.5	$D_R$ fraction	27

The active material present in the original potato extract also appeared to be hypoxanthine. Examination of an active Ca fraction showed little or no guanine or xanthine, some adenine and considerable hypoxanthine.<sup>6</sup> The quantity of hypoxanthine found in this extract was sufficient to account for the activity of the fraction. The hypoxanthine crystallized from the potato extract was active; the adenine was inactive. Furthermore, the Ca fraction was as active (or more active) after treatment with nitrous acid as before. This would be expected if the active material in the original potato extract were hypoxanthine. We are inclined to believe, therefore, that factor  $Z_1$  in potato extract is hypoxanthine though in other materials it may be guanine or a mixture of the two.

Hydrolysis of guanosine, guanylic acid, sodium guanylate, guanosine picrate, monomethyl guanosine, sodium inosine, barium inosinate and monoacetone guanosine yielded active preparations probably because of the guanine or hypoxanthine formed in the hydrolysis. Both samples of folic acid also gave active material on hydrolysis, that from Peterson being

the more active. This indicates that folic acid probably contains guanine or hypoxanthine.

The small quantity of hypoxanthine which is effective (activity has been observed for 0.3 m $\mu$  mole) is evidence for its functioning as a growth substance. Furthermore, the inactivity of xanthine, adenine and other compounds closely related to guanine or hypoxanthine demonstrates a considerable degree of specificity for the action of these two compounds and suggests that they may function in an enzyme system as thiamine, riboflavin, adenine and nicotinic acid are known to function. The inactivity of xanthine shows that the replacement of the NH<sub>2</sub> radical or hydrogen in the second position on the purine ring with oxygen renders guanine or hypoxanthine inactive (table 2). The inactivity of adenine suggests that oxy-



FIGURE 1

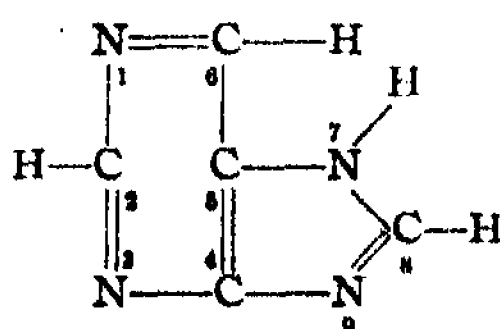
(1) Crystals of the silver salt of hypoxanthine prepared from pure hypoxanthine.  
(2) Crystals obtained with ammoniacal silver nitrate from potato extract which had been treated with nitrous acid and which showed factor Z<sub>1</sub> activity. Magnified 160  $\times$ .

gen in the sixth position is important because the substitution of NH<sub>2</sub> for oxygen in hypoxanthine renders that compound inactive. The inactivity of 2-amino uric acid suggests that the hydrogen in the eighth position is important and may not be replaced by oxygen. The inactivity of guanylic acid, guanosine and similar compounds indicates that position 7 or 9' must be open, perhaps for combination with some other compound than the sugar which is present in guanosine. The inactivity of isoguanine emphasizes the importance of oxygen in the sixth position and hydrogen or NH<sub>2</sub> in the second position. It would appear that the active substance must have the eighth position and the seventh or ninth position open for substitution, oxygen in the sixth position and hydrogen or NH<sub>2</sub> in the second position.

We can only speculate as to whether hypoxanthine (or guanine) is active *per se*, or whether it, together with an unidentified substance, makes up a larger and active molecule, as, for example, thiazole and pyrimidine combine to form thiamine. Our observations have shown that factor  $Z_1$  (hypoxanthine or guanine) is much more effective in the presence of a second unidentified substance, factor  $Z_2$ , than it is when used alone. Further information on the relation between factor  $Z_1$  and factor  $Z_2$  will have to wait the isolation of the latter substance.

TABLE 2

POSITION ON THE PURINE RING OF VARIOUS RADICALS FOR SOME OF THE COMPOUNDS TESTED FOR FACTOR  $Z_1$  ACTIVITY



SUBSTANCE	POSITION ON PURINE RING							ACTIVITY
	1	2	3	6	7	8	9	
Guanine	H	NH <sub>2</sub>	..	O	H	H	..	Active
Hypoxanthine	H	H	..	O	H	H	..	Active
Isoguanine	..	O	H	NH <sub>2</sub>	H	H	..	Inactive
Xanthine	H	O	H	O	H	H	..	Inactive
Adenine	..	H	..	NH <sub>2</sub>	H	H	..	Inactive
2-amino- uric acid	H	NH <sub>2</sub>	..	O	H	O	H	Inactive
Theobromine	H	O	CH <sub>3</sub>	O	CH <sub>3</sub>	H	..	Inactive
Theophyllin	CH <sub>3</sub>	O	CH <sub>3</sub>	O	H	H	..	Inactive
Guanosine	H	NH <sub>2</sub>	..	O	sugar	H	..	Inactive

<sup>1</sup> Robbins, W. J., *Am. Jour. Bot.*, 26, 772-778 (1939); Robbins, W. J., *Bot. Gaz.*, 101, 428-449 (1939); Robbins, W. J., *Am. Jour. Bot.*, 27, 559-564 (1940); Robbins, W. J., and Hamner, K. C., *Bot. Gaz.*, 101, 912-927 (1940); Robbins, W. J., *Ibid.*, 102, 520-535 (1940).

<sup>2</sup> Robbins, W. J., and Kavanagh, F., *Proc. Nat. Acad. Sci.*, 28, 4-8 (1942).

<sup>3</sup> The Ca fraction was the material adsorbed from an extract of potato tubers on charcoal and eluted with ammoniacal acetone. The  $D_R$  fraction was the filtrate from the charcoal treated extract.

<sup>4</sup> Theocin is a trade name for theophyllin. Our preparation of theocin was synthetic.

<sup>5</sup> The crystals of the silver salt of hypoxanthine were prepared as follows: A Ca fraction from potato tubers was treated with nitrous acid and electrolyzed. The cathode fraction was evaporated to small volume and the purines precipitated with ammoniacal silver nitrate. The precipitate was dissolved in dilute hot nitric acid. The solution was cooled, the precipitate removed and redissolved in hot dilute nitric acid. This procedure was repeated five times.

<sup>6</sup> The precipitate formed from 500 ml. of a Ca fraction by ammoniacal  $AgNO_3$  was freed from silver by treatment with HCl. Picric acid was added and the precipitate filtered off. Ammoniacal  $AgNO_3$  was added to the filtrate and the precipitate was dis-

solved in hot dilute  $\text{HNO}_3$ . The precipitation with  $\text{AgNO}_3$  and solution in  $\text{HNO}_3$  was repeated five times. The pale yellow solution was treated with a small amount of Norit A and filtered. From the filtrate 15 mg. of crystals were obtained identical in appearance with the silver salt of hypoxanthine.

The precipitate of picrates was dissolved in dilute  $\text{NH}_4\text{OH}$  and treated with  $\text{AgNO}_3$ . The precipitate was filtered off, washed, decomposed with  $\text{HCl}$  to remove the silver and filtered. The filtrate was made slightly alkaline with  $\text{NH}_4\text{OH}$  and allowed to stand for 24 hours. No guanine precipitated. It was treated with nitrous acid to convert any guanine and adenine to xanthine and hypoxanthine, respectively. Silver nitrate was added and the precipitate examined. It appeared to consist of the silver salt of hypoxanthine. In the examination of the picrate precipitate volumes were kept to less than 20 ml.

<sup>7</sup> There appears to be some uncertainty as to whether the sugar in guanosine is attached to the purine ring in the seventh or the ninth position. It is usually represented in the seventh position.

---

## SIZE GENES OF MICE

BY W. E. CASTLE

DEPARTMENT OF GENETICS, UNIVERSITY OF CALIFORNIA

Communicated January 21, 1942

In previous papers<sup>1</sup> it has been shown that several mutant genes of the house mouse, in addition to other effects which they exert, act also as size genes, either increasing or decreasing general body size.

The present paper contains a report on a study of the effects of two coat color genes, pink eye<sub>2</sub> ( $p_2$ ) and lethal yellow ( $A^y$ ) on body size, separately and in association with each other, but always on a background of homozygous brown ( $bb$ ). In a previous paper it was shown that the influence of  $p_2$  on body size is increased by association with brown. The same is probably true of  $A^y$  also, although qualitatively its action is the opposite of that of  $p_2$ , since it increases rather than decreases body size.

In order to learn more about the interaction of the three mutant genes,  $p_2$ ,  $A^y$  and  $b$ , crosses were arranged which would produce as litter mates mice of the four genetic classes (1) pink-eyed brown ( $aa\ bb\ p_2p_2$ ), (2) pink-eyed yellow ( $A^y\ a\ bb\ p_2p_2$ ), (3) brown ( $aa\ bb\ P_2\ p_2$ ) and (4) yellow ( $A^y\ a\ bb\ P_2\ p_2$ ).

All four groups would be homozygous for  $b$ . Groups (1) and (2) would be homozygous for  $p_2$  also, but groups (3) and (4) heterozygous for  $p_2$ . Groups (1) and (3) would be free from the  $A^y$  gene, but (2) and (4) would contain it. All possible combinations of  $p_2$  and  $A^y$  would thus be obtained associated uniformly with  $bb$ . The symbol for brown will accordingly be omitted from the genetic formulae hereafter.

TABLE 1

AVERAGE BODY SIZE OF MICE OF FOUR DIFFERENT GENOTYPES BORN AS LITTER MATES AND REARED TOGETHER, SEXES APART									
GENOTYPE	NO.	WEIGHT	REL. SIZE	BODY	REL. SIZE	TAIL	REL. SIZE		
(1) ♂ $aa\ p_1\ p_1$	83	29.41 ± 0.28	91.9	94.76 ± 0.22	98.4	81.40 ± 0.26	98.1		
(2) ♂ $A^va\ p_1\ p_1$	88	35.58 ± 0.51	111.2	96.87 ± 0.25	100.6	82.47 ± 0.28	99.3		
(3) ♂ $aa\ P_1\ p_1$	58	31.98 ± 0.32	100	96.26 ± 0.31	100	83.00 ± 0.41	100		
(4) ♂ $A^va\ P_1\ p_1$	68	37.40 ± 0.56	116.9	98.12 ± 0.26	101.9	85.40 ± 0.26	102.8		
(1) ♀ $aa\ p_1\ p_1$	72	24.60 ± 0.23	100.9	90.94 ± 0.33	99.8	79.19 ± 0.29	97.1		
(2) ♀ $A^va\ p_1\ p_1$	79	30.31 ± 0.39	124.3	93.05 ± 0.28	102.1	80.31 ± 0.29	98.5		
(3) ♀ $aa\ P_1\ p_1$	33	24.37 ± 0.38	100	91.12 ± 0.25	100	81.50 ± 0.52	100		
(4) ♀ $A^va\ P_1\ p_1$	78	33.67 ± 0.58	138.1	95.21 ± 0.26	104.4	83.05 ± 0.30	101.9		
All ♂ $p_1p_1$ , $\frac{(1) + (2)}{2}$		32.49	93.6	95.81	98.5	81.93	97.3		
All ♂ $P_1p_1$ , $\frac{(3) + (4)}{2}$		34.69	100	97.19	100	84.20	100		
All ♀ $p_1p_1$ , $\frac{(1) + (2)}{2}$		27.45	94.5	91.99	98.7	79.75	96.9		
All ♀ $P_1p_1$ , $\frac{(3) + (4)}{2}$		29.02	100	93.16	100	82.27	100		
Both sexes $p_1p_1$		...	94.0	...	98.6	...	97.1		
All ♂ $aa$ , $\frac{(1) + (3)}{2}$		30.69	100	95.51	100	82.20	100		
All ♂ $A^va$ , $\frac{(2) + (4)}{2}$		36.49	118.9	97.49	102.1	83.93	102.1		
All ♀ $aa$ , $\frac{(1) + (3)}{2}$		24.48	100	91.03	100	80.34	100		
All ♀ $A^va$ , $\frac{(2) + (4)}{2}$		31.99	130.6	94.13	103.4	81.68	101.6		
Both sexes $A^va$		...	124.7	...	102.7	...	101.8		



In this study, as in earlier ones, the young mice were separated as to sex at weaning, and kept until six months old, when they were weighed, chloroformed and measured as to body length and tail length.

In table 1 is shown for each sex separately the average size of the mice of each group as indicated by the three criteria, weight, body length and tail length.

Group (3) consisting of  $aa P_2 p_2$  animals shows body size influenced presumably by the brown gene alone of those under consideration, since the  $A^y$  gene is absent and the  $p_2$  gene present only as an unseen recessive, probably with little if any influence on body size. This group then should give us a standard for comparison of the effects of the genes  $p_2$  and  $A^y$  in various combinations. Average body weight of males in group (3) is 31.98 g., average body length 96.26 mm., average tail length 83.00 mm. The data for group (3) females are less satisfactory because of insufficient numbers. For the present we may pass them by.

Group (1) consisting of  $aa p_2 p_2$  animals shows body size influenced by both the brown gene and the pink eye gene in association. Average weight of males in this group is 29.41 g., average body length 94.76 mm., average tail length 81.40 mm. These are decreases (from the group (3) standard averages) of 8 per cent in weight, 1.5 per cent in body length and 2.0 per cent in tail length, which must be ascribed to the influence of  $p_2$  now homozygous. Decrease is shown by females also of group (1) as regards body length (0.2 per cent) and tail length (2.9 per cent) but not as regards weight. Less importance attaches to these differences because of the insufficient number of females in group (3).

Group (2) consisting of  $A^y a p_2 p_2$  individuals differs from group (1) only in the substitution of an  $A^y$  gene for  $a$ . A marked increase in body size results amounting in males to 20.9 per cent in weight, 2.2 per cent in body length and 1.3 per cent in tail length. In females the corresponding increases are 23.2 per cent in weight, 2.3 per cent in body length and 1.4 per cent in tail length. The  $A^y$  gene, as is well known, increases weight chiefly by leading to accumulation of fat especially in females, hence the large percentage of increase in weight.

Group (4), consisting of  $A^y a P_2 p_2$  individuals, gives us by comparison with group (3) another opportunity to measure the influence of  $A^y$  in substitution for  $a$ . In both these groups the influence of  $p_2$  is negligible, since it is present only as an unseen recessive. The differential factor is  $A^y$ .

Group (4) males show an increase over group (3) males of 16.9 per cent in weight, 1.9 per cent in body length and 2.8 per cent in tail length. For females the corresponding increases are 38.1 per cent in weight, 4.4 per cent in body length and 1.9 per cent in tail length. But these figures are less reliable because of the small number of females in group (3) as already noted.



The relative body size of mice of each of the four groups, taking group (3) animals as a standard of comparison, is shown in table 1. Also by combining the data on groups (1) and (2) we obtain average values for all animals homozygous for  $p_2$ , which may be compared with the averages for all animals heterozygous for  $p_2$ , obtained by combining the data for groups (3) and (4). These combination values are shown in the lower half of table 1.

By combining in a similar way the data for groups (1) and (3), we obtain values for all animals *free from the  $A^y$  gene* ( $aa$  animals) which may be compared with the values obtained by combining groups (2) and (4), which include all animals *possessing the  $A^y$  gene*. See the lowest section of table 1.

We thus obtain a comparison, based on all the available data, of the effectiveness of  $p_2$  in decreasing body size and of  $A^y$  in increasing it, from which it appears that the influence of  $A^y$  is greater on weight and body length but not on tail length, which, however, is least valuable of the three criteria of body size.

The net percentage changes effected by the two genes are as follows:

	IN WEIGHT	IN BODY LENGTH	IN TAIL LENGTH
by $p_2 p_2$	- 6.0%	-1.4%	-2.9%
by $A^y a$	+24.7%	+2.7%	+1.8%
Difference	+18.7%	+1.3%	-1.1%

These figures obtained by summarizing all the data furnished by the four groups agree well with those based on group (2) alone, in which both  $p_2 p_2$  and  $A^y a$  were simultaneously operative, which are for weight a net percentage change of +17.7, for body length +1.3 and for tail length -1.1.

The results obtained in this experiment are similar both qualitatively and quantitatively to those obtained from earlier experiments. They show that the gene  $p_2$  when homozygous decreases body size as judged by the three criteria, weight, body length and tail length. They show also that the gene  $A^y$  when heterozygous increases body size, notably in weight because of "adiposity," but also though in lesser degree in body length and tail length. When both influences are simultaneously in operation, the influence of  $A^y$  surpasses that of  $p_2 p_2$ , much as regards weight, slightly as regards body length, but little or none as regards tail length.

*Summary.*—The mutant gene, pink-eye<sub>2</sub>, decreases general body size in mice less than lethal yellow increases it, as regards both weight and body length, but not as regards tail length.

<sup>1</sup> See Castle, W. E., *Genetics*, 26, 177-191 (1941).

## THE OXIDATION OF 3,5-DIIODOTYROSINE TO THYROXINE\*

BY TREAT B. JOHNSON AND LYNDON B. TEWKESBURY, JR.†

DEPARTMENT OF CHEMISTRY, YALE UNIVERSITY

Communicated February 2, 1942

A physiologically active protein product from which *thyroxine* (IV) can be isolated was prepared from milk casein by W. Ludwig and P. von Mutzenbecher<sup>1</sup> in 1939, by treatment of this protein with iodine under carefully controlled conditions. The experimental technique of these workers was repeated by C. R. Harington and Rosalind V. P. Rivers<sup>2</sup> in England, and the original findings of the German investigators were confirmed in every respect. There appears to be no doubt as to the correctness of the conclusions reached by these independent workers and it, therefore, becomes a matter of immediate chemical interest to consider any possible mechanism of reaction whereby such an important biochemical transformation can theoretically be brought about.

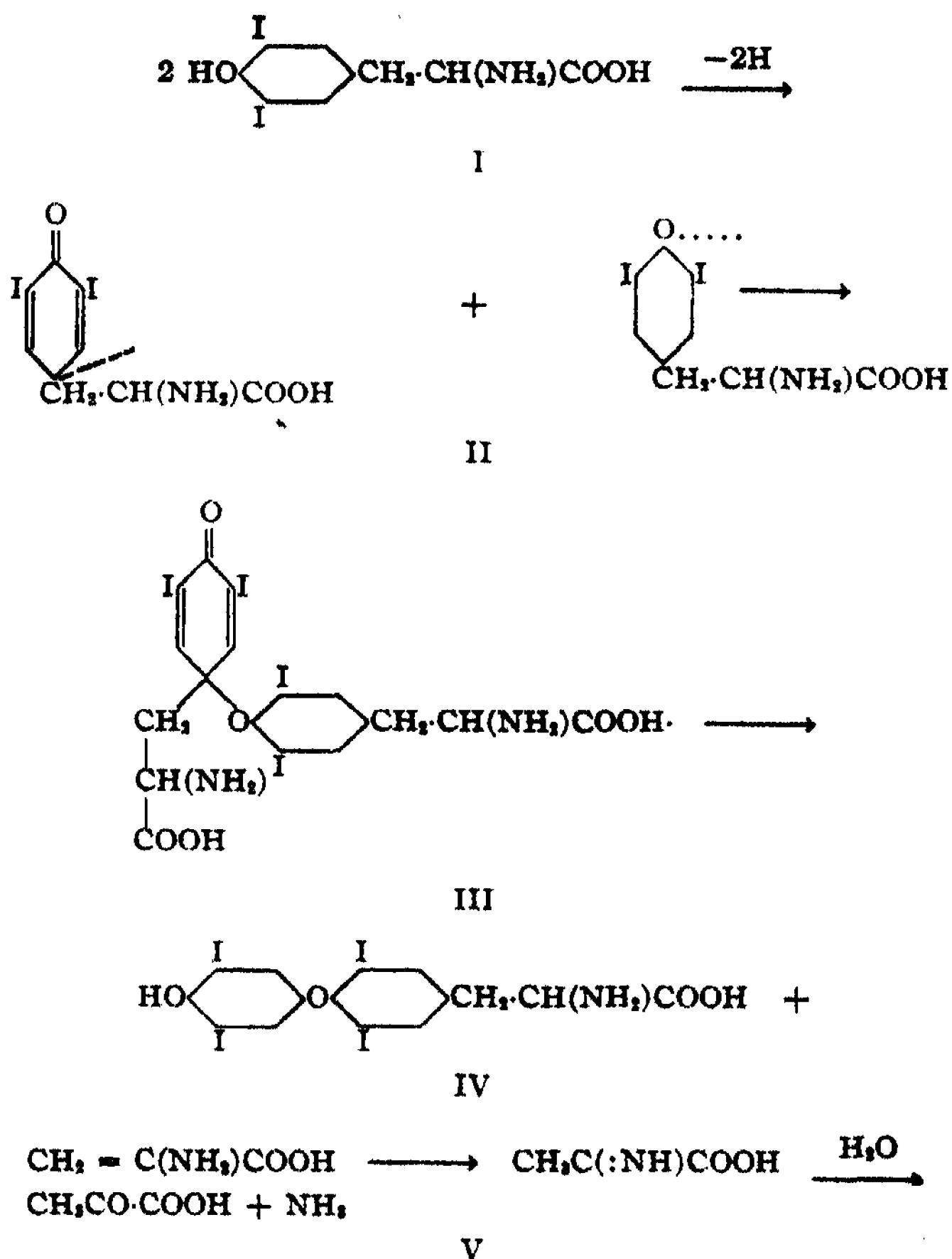
That iodine under proper experimental conditions is capable of affecting an oxidative coupling of two molecules of 3,5-diiodotyrosine to form a thyroxine molecule was established later by P. von Mutzenbecher,<sup>3</sup> who accomplished this change successfully by oxidation of 3,5-diiodotyrosine with hypiodous acid formed during prolonged incubation of this  $\alpha$ -amino acid in alkaline solution at 37°. These experiments were repeated and the results reported by P. von Mutzenbecher<sup>3</sup> were confirmed by P. Block, Jr.,<sup>4</sup> in 1940. The original hypothesis advanced by C. R. Harington<sup>2</sup> that the biogenic synthesis of *thyroxine* probably results from the iodination of tyrosine functioning in the protein molecule through the intermediate stage of 3,5-diiodotyrosine must, therefore, receive careful consideration in the light of these most interesting results.

*Biogenic Synthesis:*

The authors have repeated the incubation experiments of P. von Mutzenbecher<sup>3</sup> using synthetical 3,5-diiodotyrosine of the highest purity, and have confirmed his conclusion that thyroxine is a product of reaction under specific experimental conditions. If the postulation be correct that thyroxine is produced here as a result of oxidation of the 3,5-diiodotyrosine by hypiodous acid (HIO) formed from this  $\alpha$ -amino acid in the alkaline solution, then the yield of thyroxine should theoretically vary according to the quantity of active hypiodate present in the solution. The authors, therefore, modified P. von Mutzenbecher's technique<sup>3</sup> by adding hypiodate to the  $\alpha$ -amino acid solution in molecular proportion with the result that the yield of thyroxine was slightly increased as postulated.

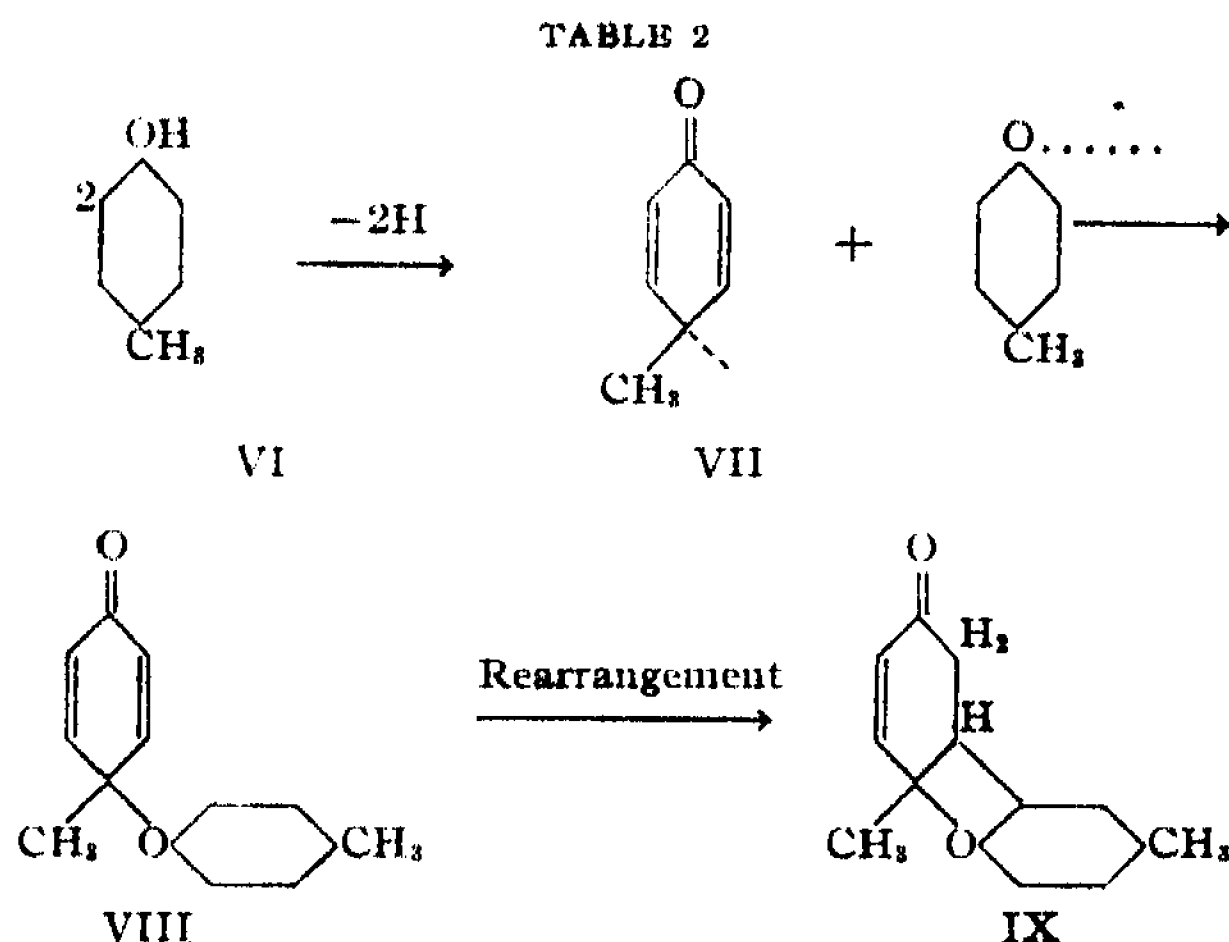
After applying several oxidation experiments with successful results, and after giving more attention to the nature of the secondary products of the reaction, the authors have been able to formulate a reaction mechanism which has not hitherto been presented to explain the formation of thyroxine (IV) from 3,5-diiodotyrosine (I). Our scheme of transformation is based on the results obtained in the fundamental studies of R. Pummerer<sup>5</sup> and co-workers dealing with the manner of oxidation of *o*- and *p*-substituted phenols in alkaline solution. We submit, therefore, the following graphical representation of this reaction (table 1) leading to the formation of thyroxine (IV) from 3,5-diiodotyrosine (I).

TABLE 1

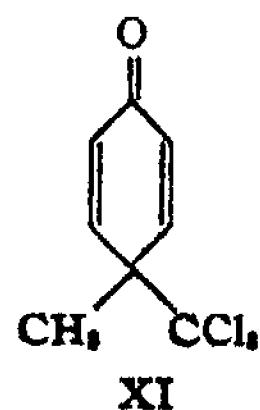
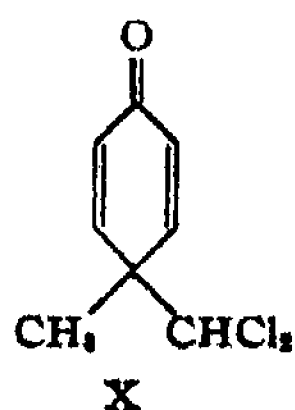


A comparison of the graphical representation of R. Pummerer's<sup>58</sup> dehydrogenation of *p*-cresol (VI) by oxidation with  $\text{K}_3\text{Fe}(\text{CN})_6$  at  $0^\circ$  (table 2)

furnishes significant data which finds application in support of the hypothesis proposed in table 1.



The chief product of this reaction is the tetrahydro-benzofuran structure expressed by formula (IX), and results apparently by rearrangement of the primary product of oxidation of *p*-cresol (VI) namely, the *p*-quinolether  $C_{14}H_{14}O_2$  (VIII). In fact, according to R. Pummerer it is characteristic of *o*- and *p*-alkylated phenols to undergo dehydrogenation easily in alkaline media to form free radicals which dimerize to more or less stable aromatic quinolethers. Such free radicals are expressed by formulae (II) and (VII), and the dimerized products by formulae (III) and (VIII) in tables 1 and 2, respectively. That *o*- and *p*-alkylated phenols may react in such a tautomeric manner to give derivatives of dihydrobenzene was also shown earlier by K. von Auwers<sup>6</sup> in 1902, who synthesized the quinone structure (X), for example, by interaction of *p*-cresol with chloroform in alkaline solution; and also by T. Zincke<sup>7</sup> who demonstrated the production of the corresponding 1-methyl-1-trichlormethyl-4-ketodihydrobenzene (XI) by application of a Friedel and Craft's reaction with *p*-cresol and carbontetrachloride.



If the hypothetical, quinoether intermediate (III), represented as formed from 3,5-diiodotyrosine (I) (table 1), is compared structurally with the corresponding oxidation product of *p*-cresol (VIII) (table 2), it may be seen that the internuclear condensation or rearrangement, which takes place in the formation of the tetrahydro-benzofurane (IX) from the *p*-quinoether (VIII), is precluded in the case of the corresponding hypothetical 3,5-diiodotyrosine derivative (III). The substitution of iodine atoms in both nuclear positions *ortho* to the ether-bound oxygen atom in (III) prevents any possibility of cyclization of this compound. Hence, stabilization of the system expressed by structure (III) must follow one of two courses namely, (1) molecular dissociation with loss of one alanine side chain and formation of thyroxine (IV) and iminopyruvic acid or, (2) hydrolysis of (III) with production of serine,  $\text{HOCH}_2\text{CH}(\text{NH}_2)\text{COOH}$  and thyroxine (IV). The results obtained by experimentation to date lead the authors to the conclusion that the formation of thyroxine (IV) from 3,5-diiodotyrosine (I) as described above is an oxidation process, and that the correctness of the reaction-mechanism postulated is supported by the facts that *pyruvic acid*<sup>8</sup> (V) and ammonia have been identified by the authors as secondary products of the reaction. Thus far, the authors have been unable to detect the presence of the  $\alpha$ -aminoacid-*serine* as a product of the oxidation reactions. A thorough search for other products of reaction is now in progress.

*Summary.*—It is a well-known fact that thyroxine and also 3,5-diiodotyrosine slowly undergo decomposition especially when exposed to sunlight, in weakly alkaline solution with formation of hypiodous acid (HIO).<sup>9</sup> This halogen acid interacts slowly with 3,5-diiodotyrosine to form thyroxine,<sup>3, 4</sup> but thus far no mechanism of reaction has been approved to explain this biochemical change. The authors have repeated the work of previous investigators of this important change, and have found that not only thyroxine but also pyruvic acid and ammonia are formed as secondary products of reaction. The discovery of *pyruvic acid* (V) as a product of this oxidation reaction has made possible the postulation of a simple and plausible mechanism to explain the formation of the *thyroxine*. The study of this problem is being continued and a discussion of the experimental technique and analytical results will be published elsewhere.

\* Work supported in part by a grant from the George E. Sheffield Research Fund of the Sheffield Scientific School of Yale University.

† Present address: 222 Summer Street, Boston, Massachusetts. Ph.D. from Yale University in June, 1941.

<sup>1</sup> Ludwig, W., and von Mutzenbecher, P., *Zeit. physiol. Chem.*, **258**, 195–211 (1939).

<sup>2</sup> Harington, C. R., and Rivers, R. V. P., *Nature*, **144**, 205 (1939).

<sup>3</sup> von Mutzenbecher, P., *Zeit. physiol. Chem.*, **261**, 253–256 (1939).

<sup>4</sup> Block, P., Jr., *Jour. Biol. Chem.*, **135**, 51–52 (1940).

- <sup>5</sup> (a) Pummerer, R., and Frankfurter, F., *Ber.*, **47**, 1472-1493 (1914).  
(b) Pummerer, R., and Cherbuliez, E., *Ibid.*, **47**, 2957-2973 (1914).  
(c) Pummerer, R., *Ibid.*, **52**, 1403-1413 (1919).  
(d) Pummerer, R., and Cherbuliez, E., *Ibid.*, **52**, 1414-1415 (1919).  
(e) Pummerer, R., and Frankfurter, F., *Ibid.*, **52**, 1416-1420 (1919).  
(f) Pummerer, R., Melamid, D., and Puttfarcken, H., *Ibid.*, **55**, 3116-3122 (1922).  
(g) Pummerer, R., Puttfarcken, H., and Schopflocher, P., *Ibid.*, **58**, 1808-1820 (1925).  
(h) Pummerer, R., and Rieche, A., *Ibid.*, **59**, 2161-2175 (1926).  
<sup>6</sup> von Auwers, K., and Winternitz, F., *Ibid.*, **35**, 465-471 (1902); von Auwers, K., and Keil, G., *Ibid.*, **35**, 4207-4217 (1902).  
<sup>7</sup> Zincke, T., and Suhl, R. S., *Ibid.*, **39**, 4148-4153 (1906).  
<sup>8</sup> The pyruvic acid is easily identified by the successful application of A. Baeyer's reaction involving the action of *o*-nitrobenzaldehyde on pyruvic acid which leads to the production of *indigo* (Baeyer, A., *Ibid.*, **15**, 2856 (1882); for other applications of this indigo color test see Johnson, T. B., and Baudisch, O., *Jour. Amer. Chem. Soc.*, **43**, 2670-2675 (1921); *Ber.*, **55**, 18-21 (1922).  
<sup>9</sup> Kendall and Osterberg, *Jour. Biol. Chem.*, **39**, 125-129 (1919).

## COMPLEMENT FIXATION WITH SIMPLE SUBSTANCES CONTAINING TWO OR MORE HAPTENIC GROUPS

BY DAVID PRESSMAN, DAN H. CAMPBELL AND LINUS PAULING

GATES AND CRELLIN LABORATORIES OF CHEMISTRY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated February 13, 1942

It has been shown that simple compounds containing two or more haptenic groups undergo typical precipitin reactions with the antibody to a protein antigen containing homologous haptenic groups.<sup>1</sup> We have investigated the question of whether or not complement fixation takes place in the formation of precipitates with these simple compounds and have obtained the results reported below.

The compounds studied are listed in table 1. Phenylarsonic acid was included, although it does not form a precipitate with antiserum, because it is known to combine with antibody and thus to inhibit the formation of precipitate by the antibody and an antigen containing two or more benzene arsonic acid groups.



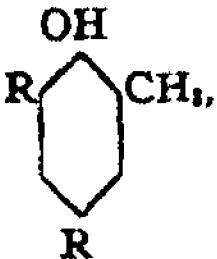
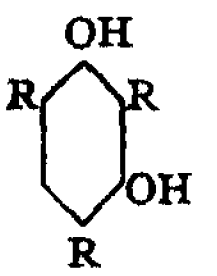

*Experimental.*—The substances I to IV and the rabbit antiserum homologous to azo sheep serum made with arsanilic acid were prepared by methods described elsewhere.<sup>16</sup> The complement of the antiserum was inactivated by heating at 57° for 30 minutes.

The amboceptor was prepared (by Mr. Carol Ikeda) by giving rabbits ten injections of 1 or 2 ml. of a 10% sheep cell suspension over a period of 3 weeks. The rabbits were bled on the seventh, eighth and ninth days after

the last injection. Its complement was inactivated by heating at 57° for 30 minutes.

TABLE 1

## COMPOUNDS INVESTIGATED

- I.  $\text{H}_2\text{O}_2\text{As}$    $\text{NHCOCONH}$    $\text{AsO}_2\text{H}_2$ , oxanilide-*p,p'*-diarsonic acid
- II. , 2-methyl-4,6-di-(*p*-azophenylarsonic acid)phenol
- III. , 1,3-dihydroxy,2,4,6-tri-(*p*-azophenylarsonic acid)benzene
- IV.   $\text{AsO}_2\text{H}_2$ , phenylarsonic acid

R represents  $\text{—NN}$    $\text{AsO}_2\text{H}_2$

Complement serum was pooled from the blood of four guinea pigs.

Complement fixation tests along with various controls were carried out by mixing saline solution, guinea pig complement (1/10 dilution) and hap-  
tene or antigen solution (20  $\mu\text{g./ml.}$ , pH 7.0) as shown in table 2. The  
mixtures were incubated for 30 minutes at 37° and the precipitates were

TABLE 2

## PROTOCOL FOR COMPLEMENT FIXATION TESTS

TEST SOLUTION	COMPLEMENT 1/10 DILUTION ML.	ANTIAZO BENZENE <i>p</i> -ARSONIC ACID SERUM ML.		SALINE ML.	ANTIGEN OR HAPTENE (20 $\mu\text{g./ML.}$ ) ML.	
1	1.5	0		3	0	Control on complement
2	1.5	1.5		1.5	0	Control on anticomplemen- tary action of antiserum
3	1.5	0		1.5	1.5 I	Control on anticom- plementary action of substances I to IV
4	1.5	0		1.5	1.5 II	
5	1.5	0		1.5	1.5 III	
6	1.5	0		1.5	1.5 IV	
7	1.5	1.5		0	1.5 I	Complement fixation tests of substances I to IV
8	1.5	1.5		0	1.5 II	
9	1.5	1.5		0	1.5 III	
10	1.5	1.5		0	1.5 IV	

removed from tubes 7, 8 and 9. To 0.5 ml. portions of each solution there were added 0.5 ml. portions of mixtures of equal volumes of 5% sheep cells and amboceptor of dilutions 1/25, 1/50, 1/100 and 1/200. The tubes were then incubated for 15 minutes at 37°. The results of the tests are shown in table 3.

TABLE 3  
RESULTS OF COMPLEMENT FIXATION TESTS  
Concentration of Amboceptor

TEST SOLUTION	1/25	1/50	1/100	1/200
1	4+	4+	+	—
2	4+	3+	+	—
3	4+	3+	+	—
4	4+	3+	2+	+
5	4+	3+	2+	+
6	4+	3+	2+	+
7	2+	—	—	—
8	2+	—	—	—
9	2+	—	—	—
10	4+	3+	+	—

4 + means complete lysis

— means no lysis

*Discussion and Summary.*—The experimental results show that complement is removed from solution by the formation of a precipitate by a multihaptenic simple substance and its haptene-homologous antiserum. A simple univalent haptene, phenylarsonic acid, was found not to fix complement under the conditions investigated.

It is accordingly not necessary for fixation of complement that the antigen be a protein or other very complex molecule; but it is indicated that combination with antibody (as by haptene) without precipitation may not be sufficient for complement fixation.

<sup>1</sup> (a) Landsteiner, K., and van der Scheer, J., *Proc. Soc. Exp. Biol. and Med.*, 29, 747 (1932); *Jour. Exp. Med.*, 56, 399 (1932); (b) Pauling, L., Campbell, D. H., and Pressman, D., *Proc. Nat. Acad. Sci.*, 27, 125 (1941); (c) Pauling, L., Pressman, D., Campbell, D. H., Ikeda, C., and Ikawa, M., *Jour. Am. Chem. Soc.*, to be published.



# MAXIMAL SYLOW SUBGROUPS OF A GIVEN GROUP

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated January 19, 1942

Suppose that a given solvable group  $G$  contains maximal Sylow subgroups of at least three different orders. Since  $G$  is solvable it contains an invariant subgroup of prime index  $p$ . This subgroup involves all the Sylow subgroups of  $G$  except those whose orders are powers of  $p$ . As the former Sylow subgroups are conjugate under this invariant subgroup as well as under  $G$  each of them is transformed into itself under  $G$  by  $p$  times as many operators as under this invariant subgroup. Hence none of these Sylow subgroups can be maximal under  $G$  unless it is of index  $p$  under  $G$ . That is, this invariant subgroup cannot contain maximal Sylow subgroups of  $G$  of more than one order. Hence *a solvable group cannot contain maximal Sylow subgroups of more than two different orders and if it contains maximal Sylow subgroups of two different orders at least one of them is invariant and of prime index under the group.*

If a group  $G$  contains maximal Sylow subgroups of exactly two different orders and two of them are invariant then they have only the identity in common and  $G$  is their direct product. That is,  $G$  is then the cyclic group of order  $pq$ ,  $p$  and  $q$  being distinct prime numbers. If only one of these Sylow subgroups is invariant under  $G$  then it must be abelian since otherwise its commutator subgroup would be transformed into itself by an operator which would be prime to its order and  $G$  would involve maximal subgroups of more than two different orders according to the theorem at the close of the preceding paragraph. For similar reasons the given invariant maximal subgroup must then be of type  $1^m$  and must admit an automorphism of prime order which does not transform any of its proper subgroups into itself. This condition is obviously sufficient as well as necessary.

From what precedes it results that a solvable group contains either no maximal Sylow subgroup or all its maximal Sylow subgroups are either of the same order or of two different orders. The last of these three cases is the simplest and was considered above. When all the maximal Sylow subgroups of a solvable group are of the same order and the group contains only one such subgroup this subgroup must be invariant and hence it must be of prime index under the group. It must therefore involve an operator whose order is equal to this index and transforms into itself at least one of the proper subgroups of the given invariant subgroup since otherwise the group would contain maximal Sylow subgroups of two different orders. The two abelian groups of order 12 are illustrations of solvable groups in which

all the maximal Sylow subgroups are of the same order and invariant, while in the dihedral and the dicyclic groups of this order the maximal Sylow subgroups are of the same order but non-invariant.

When every maximal subgroup of a given group, which may not be solvable, is a Sylow subgroup of  $G$  and  $G$  contains only two maximal subgroups then each of these subgroups is invariant under  $G$  and  $G$  is their direct product since they could not have any common operator besides the identity. Moreover, each of these two maximal subgroups is of prime order since otherwise  $G$  would involve a maximal subgroup which would not be a Sylow subgroup. That is, when every maximal subgroup of  $G$  is a Sylow subgroup and  $G$  contains only two maximal subgroups then  $G$  is the cyclic group of order  $pq$ ,  $p$  and  $q$  being distinct prime numbers. When every maximal subgroup of  $G$  is a Sylow subgroup and  $G$  contains more than two maximal subgroups then it is not possible for all of these subgroups to be invariant under  $G$  since otherwise  $G$  would contain a maximal subgroup which would not be a Sylow subgroup. Hence at least one of these maximal subgroups appears in a complete set of conjugate subgroups of  $G$  involving more than one subgroup, and since all the Sylow subgroups of the same order contained in  $G$  appear in the same set of conjugates under  $G$  there is one and only one set of conjugate subgroups of  $G$  which involves this particular Sylow subgroup.

The given set of  $n$  conjugate subgroups is transformed under  $G$  according to a primitive permutation group since the subgroup composed of all the permutations of this group which omit a given letter is maximal. This primitive permutation group satisfies the condition that each of its subgroups composed of all its permutations which omit a given letter is of degree  $n - 1$  and that all of its permutations besides the identity are of degree  $n - 1$ . For if two such subgroups had more than the identity in common then these common operators would be transformed into themselves by operators of each of these subgroups which are not common to the two subgroups since a proper subgroup of a prime power group is always transformed into itself by operators of this group which do not appear in this subgroup. It therefore would result that  $G$  would contain a maximal subgroup which would not be a Sylow subgroup since the proper subgroup composed of all the operators of  $G$  which would transform into themselves these common operators of the two given subgroups would either be a maximal subgroup of  $G$  or it could be extended by operators of  $G$  so as to obtain a maximal subgroup of  $G$ .

As this is impossible the given primitive group of degree  $n$  is of class  $n - 1$  and hence it contains an invariant regular subgroup of order  $n$  according to a well-known theorem due to G. Frobenius.<sup>1</sup> This primitive permutation group is simply isomorphic with  $G$  since all the maximal subgroups of  $G$  are supposed to be Sylow subgroups of  $G$ . For the same reason the given in-

variant subgroup of order  $n$  is abelian and has an order which is a power of a prime number. It is of type  $1^m$  since otherwise it would contain a characteristic proper subgroup and hence  $G$  would contain a maximal subgroup which would not be a Sylow subgroup. This proves the following theorem: *If every maximal subgroup of the group  $G$  is a Sylow subgroup then the order of  $G$  is divisible by two and only two distinct prime numbers and there is at least one invariant maximal subgroup in  $G$ . If  $G$  contains more than one invariant maximal subgroup it contains exactly two such subgroups and is the cyclic group of order  $pq$ ,  $p$  and  $q$  being distinct prime numbers. If  $G$  contains only one invariant maximal subgroup this is abelian, of type  $1^m$ , and of prime index under  $G$ .*

From this theorem it follows that when every maximal subgroup of a given group is a Sylow subgroup thereof then each of its Sylow subgroups is also a maximal subgroup. We proceed to consider the case when every Sylow subgroup of a given group is a maximal subgroup thereof. If such a group contains a maximal subgroup which is not also a Sylow subgroup this maximal subgroup cannot contain a Sylow subgroup of the group since this Sylow subgroup is supposed to be maximal. Its index under the group would therefore be divisible by each of the prime numbers which divide the order of the group. In particular, it could not be invariant under the group since a maximal invariant subgroup of a group is of prime index under the group. That is, if every Sylow subgroup of a group is a maximal subgroup of the group then the group cannot contain an invariant maximal subgroup unless it is also a Sylow subgroup, and if it is a Sylow subgroup it is of prime index under the group and comes under the theorem noted in the preceding paragraph.

Suppose that the group  $G$  contains maximal Sylow subgroups of order  $p^2$ ,  $p$  being a prime number. If two of these subgroups have only the identity in common then every two of them have only the identity in common, and hence they are then transformed under  $G$  according to a primitive permutation group which contains a regular invariant subgroup whose order is equal to the number of these subgroups of order  $p^2$  and hence its order would be prime to  $p$ . As the corresponding quotient group would be of order  $p^2$  it would not be possible for every Sylow subgroup of  $G$  to be maximal. If two of the maximal subgroups of order  $p^2$  would have a subgroup of order  $p$  in common this subgroup would be invariant under  $G$  and to a Sylow subgroup of the corresponding quotient group whose order would be prime to  $p$  there would correspond a subgroup of  $G$  which would not be a Sylow subgroup of  $G$  but would be either a maximal subgroup of  $G$  or could be extended to such a subgroup.

Hence it results that if every Sylow subgroup of  $G$  is maximal and  $G$  contains a Sylow subgroup of order  $p^2$  then it contains only one such Sylow subgroup and hence is of order  $p^2q$ ,  $p$  and  $q$  being distinct prime numbers,

and it contains an invariant subgroup of order  $p^2$ . This invariant subgroup is non-cyclic and none of its subgroups of order  $p$  is transformed into itself by an operator of order  $q$ . The tetrahedral group and a group of order 75 which involves the non-cyclic group of order 25 are instances of such a group. On the other hand when the given invariant subgroup is the non-cyclic group of order 9 operators of orders 2 and 3 in the group of isomorphisms would transform into itself a subgroup of order 3 contained in this invariant subgroup and hence not every Sylow subgroup would be maximal. It therefore results that *if every Sylow subgroup of a group is maximal and the group contains maximal subgroups which are not Sylow subgroups then the group must not only be insoluble but each of its Sylow subgroups must have an order which is at least the cube of a prime number.*

<sup>1</sup> Cf. Miller, Blichfeldt, Dickson, *Finite Groups*, second edition, p. 113 (1938).

## AN INVERSE PROBLEM CONCERNING A CHAIN PROCESS

BY I. OPATOWSKI

INSTITUTE OF TECHNOLOGY, UNIVERSITY OF MINNESOTA

Communicated January 19, 1942

The processes in which one or more entities change their characteristics in a *successive manner* are *chain processes*. They have been studied in a deterministic and in a statistic manner.

In the *deterministic* treatment the entities involved in the process are grouped into a number of different kinds: {kind 0, kind 1, ..., kind  $n$ }. Let  $N_i(t)$  be the amount of the kind  $i$  at the time  $t$ . If at the moment  $t = 0$ , in which the process started, all entities were of kind 0, then:

$$Y_0(0) = 1, Y_i(0) = 0 \text{ for } i \geq 1, \text{ where } Y_i(t) = N_i(t)/N_0(0). \quad (1)$$

If each kind  $i$  changes directly only into the kind  $i + 1$  and if the amount of the kind  $i$  changed during  $dt$  into the kind  $i + 1$  is proportional to  $N_i(t)$ , then:

$$dY_i/dt = k_i Y_{i-1}(t) - k_{i+1} Y_i(t), \quad (i = 0, 1, \dots, n), \quad (2)$$

where  $k_i$ 's are constants with  $k_0 = 0$ . If  $r_\rho$  constants  $k_i$ 's are equal to each other, the symbol  $K_\rho$  will be used instead of  $k_i$  ( $\rho = 1, \dots, m$ ;  $\sum_\rho r_\rho = n + 1$ ).

The system (2) as a mathematical expression of a chain process was introduced by Rutherford<sup>1</sup> to study the radioactive transformations, and

has been applied since then in many other fields. In chemical kinetics it represents a general type of successive reactions.<sup>2</sup> In the theory of biological effects of radiations it gives a relation between the number of killed microorganisms and the time of irradiation.<sup>3</sup> In the theory of photographic emulsions<sup>4</sup> it explains the formation of a latent image in relation to the time of exposure.

In the *statistic* treatment the process is a particular Markoff chain. All entities involved in the process are considered as forming a system which may take different states: {state 0, state 1, ..., state  $n$ }. Let  $p_{ij}(t)$  be the probability of being the system at the time  $T + t$  in the state  $j$ , if it is at the time  $T$  in the state  $i$ . If the process is *homogeneous with respect to the time*, i.e., if  $p_{ij}(t)$  is independent of  $T$  and if  $p_{ij}(t)$  is continuous on the right at  $t = 0$ , then by a theorem of W. Doeblin<sup>5</sup>  $p_{ij}(t)$  is necessarily a differentiable function when  $t > 0$  and is differentiable on the right at  $t = 0$ . We consider a process of this type and restrict ourselves to the case in which an arbitrary state  $i - 1$  may change during  $dt$  only into the state  $i$ , and the probability for this is  $k_i dt$ , where  $k_i$  (*an intensity*<sup>6</sup> of the process) is a constant. Then<sup>7,8,9</sup> (1) and (2) are satisfied by  $p_{0,i}(t) \equiv Y_i(t)$  and also by the other probabilities  $p_{ij}(t)$ .

This statistic form of the process was first considered by H. Bateman<sup>9</sup> in connection with bombardment of a fluorescent screen by  $\alpha$  particles. It occurs in general in radioactive disintegrations and was applied by W. Feller<sup>10</sup> to the growth of biological populations. Into the same form may be included also a chain process whose intensities depend on the time and are functions of the type:  $k_i(t) = f(i)\varphi(t)$ , because it reduces to a process with constant intensities by using a suitable time variable (of such a type is for instance the Pólya stochastic process<sup>6</sup>).

In many applications the only kinds or states of the changing system accessible to quantitative measurements are the final ones. For instance in killing microorganisms by irradiation, the various degrees of injury which the organisms undergo before dying, are usually not distinguishable with an accuracy necessary for a numerical counting, whereas dead organisms are clearly identified. The same situation occurs in radioactive transformations when the intermediate products are unstable or when their radiation energy is too small for measurements.<sup>11</sup> In such cases  $Y_n(t)$  is the only  $Y_i$ -function which is obtained from the experiments and the problem arises to determine the laws governing the process, i.e., to evaluate the constants  $k_i$ 's and the number  $n$  knowing  $Y_n(t)$ . This problem is the object of the present paper.<sup>12</sup> Essential tools for its solutions are the *Laplace transformation*, which Bateman<sup>13</sup> used to solve (2), and the *complete homogeneous symmetric functions*<sup>14</sup> defined by:  $h_0 = 1$ ;  $h_\nu(k_1, \dots, k_{n+1}) = \sum \Pi_i^{\alpha_i} k_i^{\alpha_i}$ , where  $\sum$  is taken for all sets of non-negative integers  $\alpha_i$  such that  $\sum_{i=1}^{n+1} \alpha_i = \nu \geq 1$ . Since this does not involve any compli-

cation we will take  $k_i$ 's as complex numbers. We put also  $k_i \neq 0$  for  $1 \leq i \leq n$ , because<sup>6</sup> if a  $k_i = 0$ , all  $Y_j \equiv 0$  for  $j \geq i$ . For  $|z| < |q_\rho^{-1}|$ , the expansion:<sup>14</sup>

$$\Pi_\rho^\rho = 1^m (1 - q_\rho z)^{-r_\rho} = \sum_{\nu=0}^{\infty} h_\nu(q_\rho | r_\rho; \rho = 1, \dots, m) z^\nu, \quad (3)$$

holds, where  $h_\nu(q_\rho | r_\rho; \rho = 1, \dots, m)$  means that each  $q_\rho$  is taken  $r_\rho$  times in calculating  $h_\nu$ . We need an expansion in partial fractions:

$$\Pi_\rho^\rho = 1^m (s + K_\rho)^{-r_\rho} = \sum_{\mu=0}^{r_\rho-1} \sum_{\rho=1}^m C_{\rho\mu} (s + K_\rho)^{\mu-r_\rho}. \quad (4)$$

Introducing here  $s = z - K_\tau$ ,  $K_\tau - K_\rho = q_\rho^{-1}$  for  $\rho \neq \tau$ , multiplying by  $z^{r_\tau}$  and taking into account (3) we get:  $\Pi_\rho^\rho = 1^m, \rho \neq \tau (-q_\rho)^{r_\rho} (1 - q_\rho z)^{-r_\rho} = \sum_{\mu=0}^{r_\tau-1} C_{\tau\mu} z^\mu + \text{terms in } z \text{ of degree } \geq r_\tau$ . Therefore, by (3):

$$C_{\tau\mu} = h_\mu \Pi_\rho^\rho = 1^m, \rho \neq \tau (K_\rho - K_\tau)^{-r_\rho}$$

where  $h_\mu$  stands for

$$h_\mu((K_\tau - K_\rho)^{-1} | r_\rho; \rho = 1, \dots, m; \rho \neq \tau).$$

The Laplace transform<sup>15</sup> of  $Y_n(t)$ , i.e.,  $y_n(s) \equiv \mathfrak{L}\{Y_n(t), s\} \equiv \mathfrak{L}\{Y_n\}$  is:<sup>18</sup>

$$y_n(s) \equiv \int_0^\infty e^{-st} Y_n(t) dt = P_n \Pi_\rho^\rho = 1^m (s + K_\rho)^{-r_\rho},$$

where  $P_n = k_1 k_2 \dots k_n$ . (5)

Therefore by (4), since<sup>15</sup>  $v!(s + K)^{-v-1} = \mathfrak{L}\{t^v e^{-Kt}\}$ :

$$Y_n(t) = P_n \sum_{\tau=1}^m \sum_{\mu=0}^{r_\tau-1} A_{\tau\mu} t^{r_\tau-\mu-1} e^{-K_\tau t},$$

where  $A_{\tau\mu} = C_{\tau\mu}/(r_\tau - \mu - 1)!$  (6)

If all  $r_\rho = 1$ , then  $A_{\tau\mu} = A_{\tau,0} = C_{\tau\mu} = C_{\tau,0} = \Pi_{i=1, i \neq \tau}^n (k_i - k_\tau)^{-1}$ , by a formula of Bateman.<sup>13</sup> It is seen from (6) that<sup>15</sup>

$$\mathfrak{L}\{Y_n(t), s\} \text{ exists for } \Re s > -\Re K_\rho, \rho = 1, \dots, m. \quad (7)$$

$Y_n(t)$  may be written for  $n \geq 1$  as a repeated convolution:<sup>15</sup>

$$Y_n(t) = P_n \exp(-k_1 t) * \exp(-k_2 t) * \dots * \exp(-k_{n+1} t), \quad (8)$$

which is equivalent to a repeated integral, because  $e^{-kt} * F(t) = e^{-kt} \int_0^t e^{kt} F(t) dt$ . It is seen from (6) or (8) that  $Y_n(t)$  is a symmetric function of  $k_1, \dots, k_n$ . Putting in (3)  $q_\rho = -K_\rho$ ,  $z = 1/s$ , we get:  $s^{n+1} \Pi_\rho^\rho = 1^m (s + K_\rho)^{-r_\rho} = \sum_{\nu=0}^{\infty} (-1)^\nu h_\nu(K_\rho | r_\rho; \rho = 1, \dots, m) s^{-\nu}$  for  $|s| > |K_\rho|$ . Therefore by (5):  $y_n(s) = P_n \sum_{\nu=0}^{\infty} (-1)^\nu h_\nu(k_1, \dots, k_{n+1}) s^{-\nu-n-1}$ , where  $h_\nu(k_1, \dots, k_{n+1}) \equiv h_\nu(K_\rho | r_\rho; \rho = 1, \dots, m)$ , and by known theorems,<sup>15</sup> since  $\mathfrak{L}\{t^\alpha\} = \alpha! s^{-\alpha-1}$ :

$$Y_n(t) = P_n \sum_{\nu=0}^{\infty} (-1)^\nu h_\nu(k_1, \dots, k_{n+1}) t^{\nu+n}/(\nu+n)!, \quad (9)$$

which<sup>15</sup> by (6) or (8) is convergent for any finite  $t$  and  $k_i$ .

We give now three methods for the solution of our problem: the determination of the  $k_i$ 's and of the  $n$  from  $Y_n(t)$ . The question what method is more convenient in a practical application depends on how  $Y_n(t)$  is known. If it is known with a very good accuracy for small values of  $t$  method I may be chosen. Method III requires an accurate knowledge of  $Y_n(t)$  for large values of  $t$  if  $n$  is large.

*Method I.* We have from (9):  $d^\kappa Y_n(0)/dt^\kappa \equiv Y_n^{(\kappa)}(0) = 0$  for  $\kappa < n$ ,  $Y_n^{(n+\nu)}(0) = (-1)^\nu P_n h_\nu(k_1, \dots, k_{n+1})$  for  $\nu \geq 0$ . Since  $h_0 = 1$ ,  $n$  is obtained as the order of the first derivative of  $Y_n(t)$  which is  $\neq 0$  at  $t = 0$ . Besides this:

$$h_\nu(k_1, \dots, k_{n+1}) = (-1)^\nu Y_n^{(n+\nu)}(0)/Y_n^{(n)}(0). \quad (10)$$

Therefore  $k_1, \dots, k_{n+1}$  are determined as the roots of the equation

$$x^{n+1} + a_1 x^n + \dots + a_{n+1} = 0, \quad (11)$$

where  $(-1)^i a_i$  are the elementary symmetric functions of  $k_1, \dots, k_{n+1}$  and may be calculated by means of the equations<sup>14</sup>  $\sum_{i=0}^n (-1)^i h_i a_n = 0$ ,  $a_0 = h_0 = 1$ .

*Method II.* This method and the following one hold if  $\Re k_i > 0$  for  $i \leq n$  and  $k_{n+1} = 0$ , or if

$$\Re k_i > 0 \text{ for } i \leq n+1. \quad (12)$$

(These restrictions on  $k_i$ 's are normally satisfied in all applications.) We assume first that (12) holds. Then by (6):  $Y_n^{(\kappa)}(\infty) = 0$  for  $\kappa \geq 0$  and

$$\int_0^\infty Y_n^{(\kappa)}(t) dt = -Y_n^{(\kappa-1)}(0) \text{ for } \kappa \geq 1. \quad (13)$$

Therefore  $n+1$  is the smallest value of  $\kappa > 0$ , for which the integral (13) is  $\neq 0$ , and by (10):

$$h_\nu(k_1, \dots, k_{n+1}) = (-1)^\nu \int_0^\infty Y_n^{(n+\nu+1)}(t) dt / \int_0^\infty Y_n^{(n+1)}(t) dt. \quad (14)$$

When  $h_\nu$ 's are known,  $k_i$ 's are calculated as in the method I. It is possible to lower the degree of equation (11) by one. In fact, by (7), (12) and (5):

$$\mathfrak{L}\{Y_n(t), 0\} = y_n(0) = \int_0^\infty Y_n(t) dt = 1/k_{n+1}, \quad (15)$$

which gives  $k_{n+1}$ . Then<sup>14</sup>

$$h_\nu(k_1, \dots, k_n) \equiv h_\nu(k_1, \dots, k_{n+1}) - k_{n+1} h_{\nu-1}(k_1, \dots, k_{n+1})$$

may be calculated by means of (14) or (10) for  $\nu = 1, \dots, n$ , and  $k_1, \dots, k_n$  are determined as the roots of an equation of type (11) of degree  $n$ .

■ As soon as  $k_{n+1}$  is known from (15),  $n$  may be determined also from the formula

$$n = 1 + \lim_{t \rightarrow 0} [t d \log_e |Y_n'(t) + k_{n+1} Y_n(t)| / dt], \quad (16)$$



which holds without any restriction on  $k_i$ 's. To prove (16) write a power series for  $Y_{n-1}(t)$  by changing in (9)  $n$  into  $n-1$ . From this it is easy to see that

$$\lim_{t \rightarrow 0} [t Y_{n-1}'(t) / Y_{n-1}(t)] = n - 1. \quad (17)$$

On the other hand (2) gives for  $i = n$ :  $\log |Y_n'(t) + k_{n+1} Y_n(t)| = \log |k_n| + \log |Y_{n-1}(t)|$ . By differentiating this with respect to  $t$  and taking into account (17) we get (16).

We show now that the method remains substantially unchanged if  $k_{n+1} = 0$  and if (12) holds only for  $i \leq n$ . Of course (16) is true and all the other formulas hold too if  $n$  is changed in them into  $n-1$ . But since by (2)  $Y_n'(t) = k_n Y_{n-1}(t)$ , (13) holds also for  $Y_n^{(\kappa)}$  if  $\kappa \geq 2$ . (15) gives now  $\int_0^\infty Y_{n-1}(t) dt = 1/k_n$ . Therefore  $(\dagger) \int_0^\infty Y_n'(t) dt = k_n \int_0^\infty Y_{n-1}(t) dt = 1$  and  $n+1$  is the smallest value of  $\kappa \geq 2$  for which the integral (13) is  $\neq 0$ .  $\sum Y_i'(t) = 0$ , by (2), because  $k_{n+1} = 0$ , therefore by (1)  $\sum Y_i(t) = 1$ . On the other hand from  $(\dagger)$  and (1) we get  $Y_n(\infty) - Y_n(0) = Y_n(\infty) = 1$  for  $n > 0$ , and we conclude that the system which was entirely in the state 0 at  $t = 0$  is at  $t = \infty$  entirely in the final state  $n$ .

*Method III (Method of Moments).* We assume first that (12) holds. Then by known theorems of the Laplace transformation:<sup>15</sup>

$$y_n^{(\nu)}(0) = (-1)^\nu \mathcal{L}\{t^\nu Y_n(t), 0\} = (-1)^\nu \int_0^\infty t^\nu Y_n(t) dt \quad (18)$$

is convergent. On the other hand from (5):  $k_{n+1} y_n(s) = \Pi_i^{-1} \prod_{i=1}^{n+1} (1 + k_i^{-1} s)^{-1}$  which by (3) is  $= \sum_{\nu=0}^\infty (-1)^\nu h_\nu(k_1^{-1}, \dots, k_{n+1}^{-1}) s^\nu$  for  $|s| < |k_i^{-1}|$ . Therefore:  $k_{n+1} y_n^{(\nu)}(0) = (-1)^\nu \nu! h_\nu(k_1^{-1}, \dots, k_{n+1}^{-1})$  and by (18):

$$\nu! h_\nu(k_1^{-1}, \dots, k_{n+1}^{-1}) = k_{n+1} \int_0^\infty t^\nu Y_n(t) dt, \quad \nu = 0, 1, \dots \quad (19)$$

This reduces the determination of the  $k_i$ 's to the solution of an algebraic equation in the same way as in the method II.

If (12) holds only for  $i \leq n$  and  $k_{n+1} = 0$ , all formulae of this method hold if we change in them  $n$  into  $n-1$ . Therefore, since now  $Y_n'(t) = k_n Y_{n-1}(t)$ , (19) becomes:  $h_\nu(k_1^{-1}, \dots, k_n^{-1}) = (\nu!)^{-1} \int_0^\infty t^\nu Y_n'(t) dt$ , which determines  $k_1, \dots, k_n$  in the same way as previously.<sup>17</sup>

<sup>1</sup> Rutherford, E., Chadwick, J., Ellis, C. D., *Radiations from Radioactive Substances*, Cambridge, 1930, pp. 11-12.

<sup>2</sup> See for instance Branch, G. E. K., Calvin, M., *Theory of Organic Chemistry*, New York, 1941, p. 357.

<sup>3</sup> Crowther, I. A., *Proc. Roy. Soc. London*, B100, 401 (1926); Opatowski, I., *Bull. Amer. Math. Soc.*, 47, 704 (1941); 48, 46 (1942).

<sup>4</sup> Cf. for instance Silberstein, L., *Jour. Optical Soc.* 31, 343 (1941).

<sup>5</sup> Doeblin, W., *Bull. d. Sc. Math.*, 62, 21-32 (1938); 64, 35-37 (1940).

<sup>6</sup> Lundberg, O., *On Random Processes and Their Application to Sickness and Accident Statistics*, Uppsala, 1940, pp. 57-68.

<sup>7</sup> Kolmogoroff, A., *Math. Ann.*, 104, 428-437 (1941).



<sup>8</sup> Feller, W., *Trans. Amer. Math. Soc.*, **48**, 513 (1940)

<sup>9</sup> Bateman, H., *Phil. Mag.*, 6 s., **20**, 704-707 (1910).

<sup>10</sup> Feller, W., *Acta Biotheor.*, **5**, Part I, 15-17, 19-21 (1939).

<sup>11</sup> Meyer, S., Schweidler, E., *Radioaktivitaet*, 2nd ed., Berlin-Leipzig, 1927, p. 32.

<sup>12</sup> If  $k_i \neq k_j$ , this problem is equivalent to that of fitting a set of numerical data to the sum of exponentials  $\sum c_i \exp(-k_i t) \equiv Y_n(t)$ . In the case in which  $c_i$ 's are independent of  $k_i$ 's various fitting methods are known: Aigner, F., Flamm, L., *Phys. Ztschr.*, **13**, 1151-1155 (1912); Walsh, J. W. T., *Proc. Phys. Soc. London*, **32**, 26-30 (1919); Levy, H., *Ibid.*, **34**, 108-113 (1922). Further references in Meyer-Schweidler<sup>11</sup> p. 61; also Bernstein, F., *Ztschr. ang. Math. u. Mech.*, **7**, 441-444 (1927) and Doetsch<sup>15</sup> p. 27-28. In our problem however  $c_i$ 's are certain functions of  $k_i$ 's, and these methods are not good.

<sup>13</sup> Bateman, H., *Proc. Phil. Soc. Cambridge*, **15**, 423-427 (1910). This paper remained unknown to some recent authors who gave less simple methods for the solution of (2): Arley, N., *Mat. Tidsskr.*, 49-51 (1939); Lundberg;<sup>6</sup> Feller.<sup>8</sup>

<sup>14</sup> Murnaghan, F. D., *Theory of Group Representations*, Baltimore, 1938, pp. 108, 112. These functions are called also homogeneous product sums; cf. Littlewood, D. E., *Theory of Group Characters*, Oxford, 1940, pp. 82, 83, 88 or, alef-functions of Wronski, s. *Enc. d. math. Wiss.*, **IB3b**, 465, 459.

<sup>15</sup> Doetsch, G., *Theorie der Laplace-Transformation*, Berlin, 1937, pp. 23, 148, 157, 43-48, 61-64 and Widder, D. V., *The Laplace Transform*, Princeton, 1941.

<sup>16</sup> Only the first term of this expansion was known, s. *Handb. d. Phys.*, **22**, 190.

<sup>17</sup> The author acknowledges helpful information received in preparing this paper from Profs. J. L. Doob, L. Silberstein, J. W. Tukey.

## ON CONFIDENCE INTERVALS

BY EDWIN B. WILSON

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated December 23, 1941

In 1927 I called attention to the fact that many statements about probability are highly elliptical and illustrated the matter by the simple case of a point-binomial universe with unknown probability  $p$  and observed value  $p_0$  in some sample.<sup>1</sup> Using the admittedly rough estimate of probability based on the standard deviation one ordinarily writes

$$p_0 - \lambda \sqrt{p_0 q_0 / n} < p < p_0 + \lambda \sqrt{p_0 q_0 / n}$$

and states that the probability that the true value  $p$  in the universe lies between the limits given may be had from a probability-integral table entered with a normal deviation of  $\lambda$  units. I urged that a better procedure would be to use for the standard deviation the value  $\sqrt{pq/n}$  obtained from the unknown  $p$  of the universe which leads to

$$\frac{p_0 + t/2}{1 + t} - \frac{\sqrt{p_0 q_0 t + t^2/4}}{1 + t} < p < \frac{p_0 + t/2}{1 + t} + \frac{\sqrt{p_0 q_0 t + t^2/4}}{1 + t} \quad (1)$$

where  $t = \lambda^2/n$ , and to the statement that if  $p$  (unknown) should be outside of those limits the probability of observing the sample would be less than that obtained from entering the probability-integral table with a normal deviation of  $\lambda$  units. I was trying to emphasize that we know nothing about the value of  $p$ , which must have whatever value it did have in the universe from which the sample was drawn, but that we could set limits based on probability calculations such that if  $p$  lay between them the chance

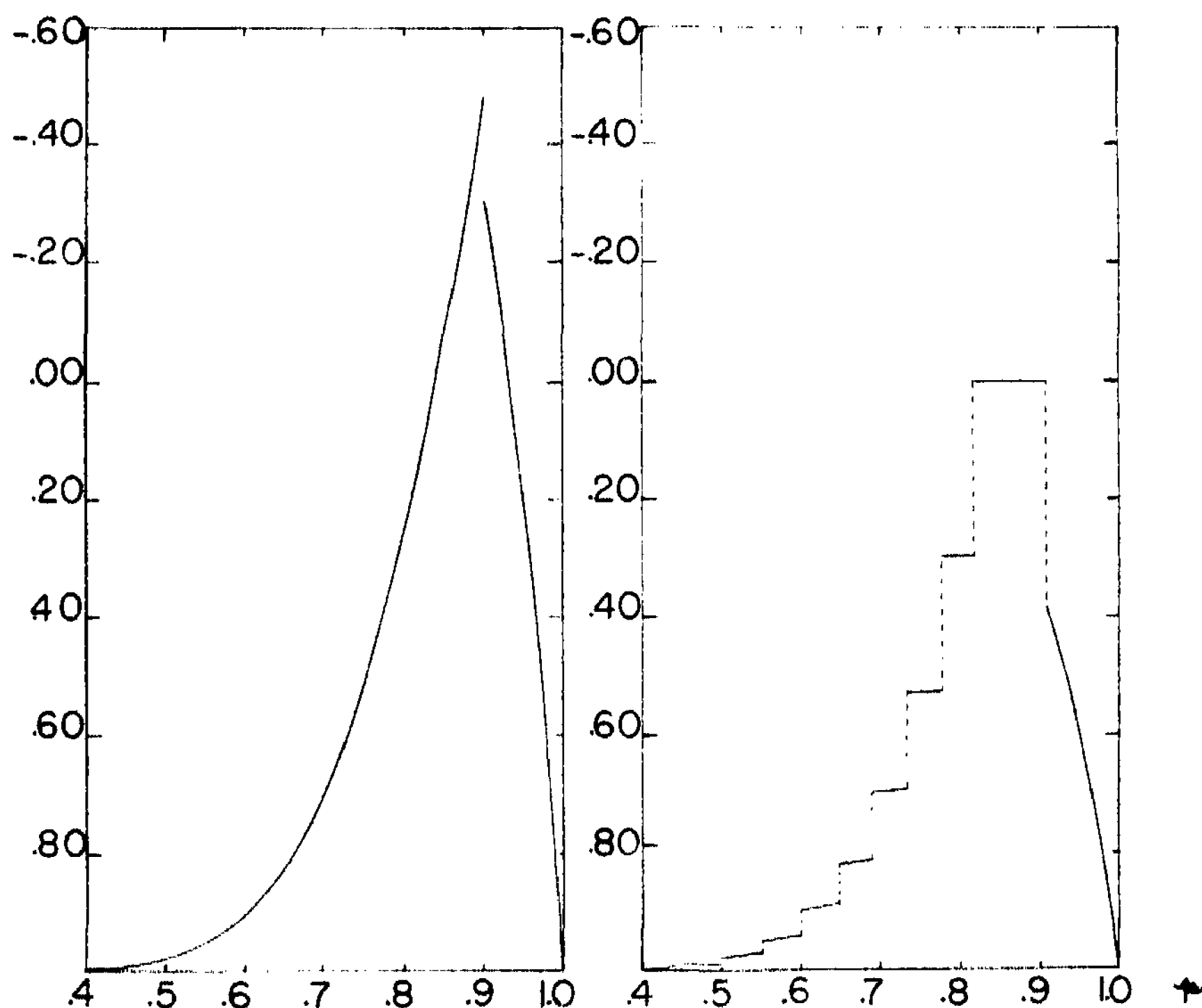


FIGURE 1

Confidence levels in terms of  $1 - P = 1 - 2\epsilon$  for  $n = 10$ ,  $x = 9$ ,  $p_0 = 0.9$  as functions of the unknown probability  $p$  on two assumptions:

- 1 (a), left: The no less extreme observation (Clopper and Pearson),
- 2 (b), right: The no more probable observation (Wilson).

of getting the particular observation or any less probable one would exceed some preassigned value  $P$  whereas if  $p$  lay outside them the chance of getting the observation or any less probable one would be less than  $P$ .

In 1934 Clopper and Pearson<sup>2</sup> introduced confidence intervals for what I take to be nearly the same notion that I was discussing. In their terminology, if  $2\epsilon$  be any positive number on which to base a certain level of con-

fidence in terms of probabilities, the interval (1) would have a confidence  $1 - 2\epsilon = 1 - P$  as judged by the test I used. Their test was different; they asked the question: For what value  $p_1$  of  $p$  less than  $p_0$  will the value  $p_0$  or a greater value arise with a probability  $\epsilon$  and for what value  $p_2$  of  $p$  greater than  $p_0$  will the value of  $p_0$  or a smaller value arise with the same probability  $\epsilon$ ? It may be noted that they divided the probability  $P = 2\epsilon$  into two equal parts and asked two questions, as though one were thinking of deviations from a central tendency, whereas I had kept to a total probability  $P$  and had asked a single question in terms of samples no more probable than the one observed.<sup>3</sup> Furthermore Clopper and Pearson (and Rietz) introduce the notion of confidence belts, which I did not have in mind, and state: "We cannot therefore say that for any specified value of  $x (= np_0)$  the probability that the confidence interval will include  $p$  is 0.95 or more. The probability must be associated with the whole belt, that is to say with the result of the continued application of a method of procedure to all values of  $x$  met with in our statistical experience"—a statement I should hesitate to make.

In figure 1 will be found for  $n = 10$  the graph of the values of  $1 - 2\epsilon = 1 - P$  against those of  $p$  for the case  $p_0 = 0.9$  or  $x = 9$  for (a) the Clopper-Pearson assumption that  $\epsilon$  is the probability of all values 9 or less or of all values 9 or greater for  $x$  according as  $p > 0.9$  or  $p < 0.9$  and for (b) my own assumption that  $P$  is the probability of the sample of 9 or of any sample no more probable than it. The figure shows that on the Clopper-Pearson basis the curve is a continuous function of  $p$  except at the value  $p = p_0 = 0.9$ ; it also shows that the shorter confidence intervals around  $p = 0.9$  have confidence values (probabilities?) less than zero.<sup>4</sup> The figure shows also that on the basis of my definition involving the probability of all no more probable cases the curve of  $P$  has multiple discontinuities.

The question of generalization may be raised.<sup>5</sup> Let there be given two independent samples of  $n$  and  $n'$  from universes supposed to have the same unknown value  $p$  but furnishing from observation two values  $p_0$  and  $p_0'$  in the two samples. If the numbers in the samples are large enough to justify estimates of probability based on values of  $\chi^2$ , one may write

$$\chi^2 = \frac{(np_0 - np)^2}{npq} + \frac{(n'p_0' - n'p)^2}{n'pq}, \quad (2)$$

$$\chi^2(p - p^2) = np_0^2 + n'p_0'^2 - 2(np_0 + n'p_0')p + (n + n')p^2,$$

which is a quadratic equation for the unknown value of  $p$ . Let

$$n + n' = N, \quad np_0^2 + n'p_0'^2 = N\bar{p}_0^2 + N\sigma_p^2, \quad t = \chi^2/N;$$

then

$$\frac{\bar{p}_0 + t/2}{1 + t} - \frac{\sqrt{\bar{p}_0 \bar{q}_0 t + t^2/4 - \sigma_p^2(1 + t)}}{1 + t} < p < \frac{\bar{p}_0 + t/2}{1 + t} + \frac{\sqrt{\bar{p}_0 \bar{q}_0 t + t^2/4 - \sigma_p^2(1 + t)}}{1 + t}. \quad (3)$$

This result is entirely similar to (1) except for the addition of the term  $-\sigma_p^2(1 + t)$  under the radical; it is, furthermore, entirely general for any number of samples drawn from different universes having the same  $p$ , provided  $\bar{p}_0$  be the weighted mean  $p_0$  for all samples and  $\sigma_p^2$  the weighted variance of the observed values  $p_0$ .

The new term  $-\sigma_p^2(1 + t)$  under the radical, being negative, shows that for a given value of  $t = \chi^2/N$  the range for  $p$  is the more restricted, the greater the scatter of the values  $p_0$  observed in the different samples; indeed we must have<sup>6</sup>

$$\sigma_p^2 \leq \frac{\bar{p}_0 \bar{q}_0 t + t^2/4}{1 + t} \quad \text{or} \quad t \geq \frac{2\sigma_p^2}{\sqrt{(\bar{p}_0 \bar{q}_0 - \sigma_p^2)^2 + \sigma_p^2} + \bar{p}_0 \bar{q}_0 - \sigma_p^2} \quad (4)$$

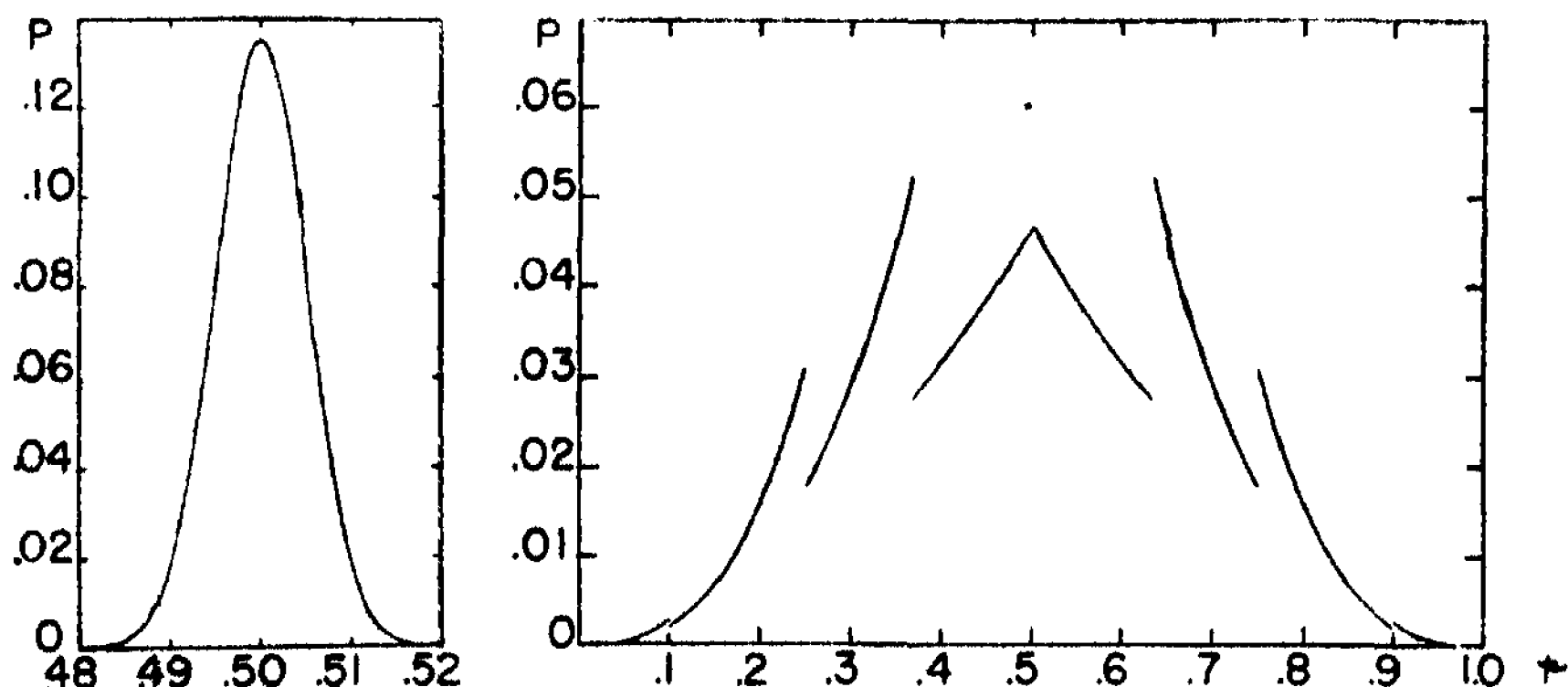


FIGURE 2

Confidence levels in terms of  $P = 2\epsilon$  as functions of the unknown probability  $p$  on the hypothesis of the no more probable observation for two cases:

2 (a), left:  $n = n' = 5000$ ,  $p_0 = 0.49$ ,  $p'_0 = 0.51$ .

2 (b), right:  $n = n' = 3$ ,  $p_0 = 1$ ,  $p'_0 = 0$ .

to have any range for  $p$ . If we have  $k$  samples and enter a  $\chi^2$  table in the row for  $k$  degrees of freedom under any value of  $P$  to find  $\chi^2$  and hence  $t = \chi^2/N$ ,  $N$  being the sum of the numbers in the  $k$  samples, we may compute from (3) the limits for  $p$  such that if  $p$  lies outside those limits the chance of getting the set of samples and all less probable sets is less than  $P$ ; if  $p$  lies within the limits (2) the chance is greater than  $P$  but is in any case less

than the value of  $P$  corresponding to that value of  $t$  or  $\chi^2$  which makes the radical in (3) vanish or corresponds to the equality sign in (4) and thus shrinks the range of  $p$  to a point.<sup>7</sup> Figure 2a gives a plot of  $P$  against  $p$  for the case  $n = n' = 5000$ ,  $N = 10,000$ ,  $p_0 = 0.49$ ,  $p_0' = 0.51$ ,  $\bar{p}_0 = 0.50$ ,  $\sigma_p^2 = 0.0001$ .

When the numbers are small the question of the test that replaces  $\chi^2$  has to be discussed. If we use the criterion of the total probability of the samples observed and of all samples of no greater probability, we may with patience discuss any particular case. For example, if  $n = 3$ ,  $p_0 = 1$ ,  $n' = 3$ ,  $p_0' = 0$ , the joint chance of the observed samples is  $p^3q^3$  whatever the unknown value of  $p$ . There are  $(n + 1)(n' + 1) = 16$  possible pairs of samples. The graph of the probability  $P$  as a function of  $p$  is in figure 2b. The graph is, of course, discontinuous. It shows that there is no value of  $p$  for which the probability  $P$  exceeds 0.0625, and this for only the single value  $p = 1/2$  which on *a priori* grounds would be infinitely rare. Values other than  $p = 1/2$  between 0.37 and 0.63 give no values greater than  $P = 0.047$ . There are very short ranges for  $p$  near 0.36 and 0.64 which give  $P > 0.05$ . Should we desire confidence intervals<sup>8</sup> corresponding to  $P = 0.05$  we should have to specify something like 0.360 to 0.366, 0.5, 0.634 to 0.640; should we desire confidence intervals corresponding to  $P = 0.01$  we should have something like  $0.17 \leq p < 0.83$ ; for  $P = 0.03$  we should have 0.248 to 0.250, 0.302 to 0.366, 0.385 to 0.615, 0.634 to 0.698, 0.750 to 0.752. The exact method of R. A. Fisher, if he would apply it to a problem of this type (as is unlikely, because the approaches seem to belong to different universes of discourse) and if in this symmetrical case he would consider together both ends of his series, would give  $P = 0.10$ ; there is no value of the unknown  $p$  which gives a probability as much as two-thirds of this amount. It thus appears that the exact method affords a different criterion from that of the total probability of the no more probable experience, just as both of these methods furnish different criteria from that based on the distribution of the difference  $p_0 - p_0'$ .

<sup>7</sup> Wilson, E. B., "Probable Inference, the Law of Succession, and Statistical Inference," *J. Amer. Statistical Association*, 22, 209-212 (1927).

<sup>8</sup> Clopper, C. J., and Pearson, E. S., "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404-413 (1934). An account of this may be found in Rietz, H. L., "A Recent Advance in Statistical Inference," *Amer. Math. Monthly*, 45, 149-158 (1938).

<sup>9</sup> There is a third notion not very different from mine or that of Clopper and Pearson—or rather of Neyman who slightly antedates them, at least with respect to continuous distributions—with which Fisher, R. A., on "Inverse Probability," *Proc. Camb. Phil. Soc.*, 26, 528-533 (1930), associated the term fiducial limits and according to him limited to continuous distributions, see "The Fiducial Argument in Statistical Inference," *Ann. Eugenics*, 6, 391-398 (1935).

<sup>4</sup> If I interpret Clopper and Pearson aright, one may logically take  $2\epsilon = 1$ ,  $\epsilon = 0.5$  and construct a confidence belt, quite as one constructs one for  $2\epsilon = 0.05$ . The belt is

of course thinner than for smaller values of  $\epsilon$  but, still, has an appreciable area within which the probability that  $p$  will lie as "the result of the continued application of a method of procedure to all values of  $x$  met with in our statistical experience" is zero. As I do not see how this can well be, I incline to some doubt in respect to their (or Rietz's) statement. Moreover, for my test as shown in figure 1 (b) there is a finite interval for which  $P = 1$  and the confidence is 0, whereas the probability that  $p$  lie in that interval could hardly be 0. Thus although confidence intervals are based on probabilities, it is not certain that probabilities are based on them. The difficulty of the negative values of probabilities  $1 - 2\epsilon$  for  $p$  lying in intervals close around  $p_0 = 0.9$  can be obviated by a modification of the definition whereby when  $p > p_0$  we use only a fraction  $\theta$  of the maximum term  $M$  for  $x = np_0$  and when  $p < p_0$  the fraction  $(1 - \theta)$  of  $M$  writing  $X + \theta M = \epsilon$  and  $(1 - \theta)M + Y = \epsilon$  where  $X$  is the sum of the terms up to but not including  $M$ , and  $Y$  is the sum of the terms beyond  $M$ , the value of  $\theta$  being  $1/2 + 1/2(Y - X)/M$ .

<sup>5</sup> I shall not try to offer methods of generalizing the Clopper-Pearson test.

<sup>6</sup> While  $\sigma_p^2$  may in extreme cases be nearly equal to  $p_0q_0$  so that  $\chi^2$  may be near  $2N\sigma_p^2 = 2N(p_0q_0)^{1/2}$ , the interesting range for  $\sigma_p^2$  when  $N$  is large is a small multiple of  $p_0q_0/N$  and thus  $\sigma_p^2$  is small compared with  $p_0q_0$ , and approximately

$$\chi^2 \cong \frac{\sigma_p^2}{p_0q_0/N} \left[ 1 + \frac{\sigma_p^2}{p_0q_0} \left( 1 - \frac{1}{4p_0q_0} \right) \right].$$

The expression in front of the bracket is the usual expression for  $\chi^2$  taken from the marginal totals of a 2 by  $k$  table and the bracket is nearly 1. Throughout this paper as in that of 1927 I have carried out formal algebraic operations as though they were significant; it should, however, be borne in mind that the expression (2) for  $\chi^2$  is discontinuous and gives from a  $\chi^2$  table the probability of the given sample or set of samples and of all no more probable ones only to a certain degree of approximation even when the numbers are large—and thus the correction in the bracket may be illusory.

<sup>7</sup> It follows that in the case now under discussion the probability  $P$  which corresponds to a certain range of  $p$  such that if  $p$  be outside it the probability of so aberrant samples is less than  $P$  does not correspond to a confidence  $1 - P$  that  $p$  lie within the interval. If  $P_0$  be the probability which shrinks the interval to a point, we might perhaps say that  $P_0 - P$  is the level of confidence that  $p$  lie within the interval and  $1 - P_0$  the level of confidence that there be no value of  $p$  common to the two universes.

<sup>8</sup> As in the previous case, the chance if  $p$  lies outside the interval is less than the specified amount, but the chance if  $p$  lies within the interval is between the specified value and some upper limit  $P_0$ , leaving  $1 - P_0$  as the level of confidence that there be no value of  $p$  common to the two samples. It may be repeated that the chances involved are not those that there actually is a common probability lying within or without certain limits but the chances that the observed samples and all no more probable should arise from such a hypothetical common value of  $p$ .

## ON CONTINGENCY TABLES

BY EDWIN B. WILSON

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated January 8, 1942

In 1934, Yates,<sup>1</sup> acting on a suggestion from Fisher, developed the so-called exact method of treating  $2 \times 2$  tables and proposed and discussed what has become known as the Yates' correction for the  $\chi^2$  expression used on such tables. Soon thereafter the method was inserted in a new edition of Fisher's Statistical Methods for Research Workers (Art. 21.02) and became rapidly and widely adopted. The method consists in considering the linear series of  $c + d + 1$  tables

$$\frac{a-d}{c+d} \Big| \frac{b+d}{0}, \dots, \frac{a-1}{c+1} \Big| \frac{b+1}{d-1}, \quad \frac{a}{c} \Big| \frac{b}{d},$$

$$\frac{a+1}{c-1} \Big| \frac{b-1}{d+1}, \dots, \frac{a+c}{0} \Big| \frac{b-c}{c+d},$$

where the convention has been adopted that  $c + d$  is the smallest marginal total, and of computing the relative probabilities

$$\frac{(a+b)!(a+c)!(b+d)!(c+d)!}{(a+b+c+d)!} \left[ \frac{1}{(a-d)!(b+d)!(c+d)!0!}, \dots, \right.$$

$$\left. \frac{1}{a!b!c!d!}, \dots, \frac{1}{(a+c)!(b-c)!0!(c+d)!} \right]$$

of those  $c + d + 1$  tables. The properties of this series of probabilities, whose sum is 1, are therefore of great importance. Connected with them is the algebraic identity

$$\sum \frac{1}{a!b!c!d!} = \frac{(a+b+c+d)!}{(a+b)!(a+c)!(b+d)!(c+d)!}$$

where the summation extends over all the  $c + d + 1$  terms of the series. The series is symmetrical when and only when two of the marginal totals on the same side of the table are equal.<sup>2</sup>

The mean value of  $a$  will be

$$\bar{a} = \frac{(a+b)!(a+c)!(b+d)!(c+d)!}{(a+b+c+d)!} \left[ \sum \frac{a}{a!b!c!d!} = \sum' \frac{1}{(a-1)!b!c!d!} \right]$$

The question is as to the value of the second sum  $\sum'$  which is extended over all tables or terms. For those tables in which  $a = 0$  the sum  $\sum$  vanishes as does the sum  $\sum'$  because  $(-1)!$  is infinite. Hence in evaluating either sum one may reject the tables for which  $a = 0$ . Consider now the series of tables arising from  $\frac{a-1}{c} \left| \frac{b}{d} \right.$  which will have the same values of  $(a-1)!b!c!d!$  as the given series. Then

$$\sum' \frac{1}{(a-1)!b!c!d!} = \frac{(a+b+c+d-1)!}{(a+b-1)!(a+c-1)!(b+d)!(c+d)!}$$

$$\text{and hence } \bar{a} = \frac{(a+b)(a+c)}{a+b+c+d} = Np_1p_2,$$

where  $N = a+b+c+d$  and where  $p_1 = (a+b)/N$ ,  $p_2 = (a+c)/N$  are the proportions observed in the margins. The proof may be extended to cover the average of any (partial) factorial as:

$$a(a-1)\dots(a-p+1) = \frac{(a+b+c+d-p)!(a+b)!(a+c)!(b+d)!(c+d)!}{(a+b+c+d)!(a+b-p)!(a+c-p)!(b+d)!(c+d)!}$$

From the values of the factorials the moments  $\mu_i$  of the distribution of  $a$  may be obtained with routine algebra. Let

$$x = a - \bar{a} = (ad - bc)/N.$$

Then<sup>3</sup>

$$\mu_2(x) = \frac{N^2}{N-1} p_1p_2q_1q_2 = \frac{(Np_1q_1)(Np_2q_2)}{N-1}$$

$$\mu_3(x) = \frac{N^3}{(N-1)(N-2)} (q_1 - p_1)(q_2 - p_2)p_1p_2q_1q_2$$

$$\kappa = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\sqrt{N-1}}{N-2} \frac{(q_1 - p_1)(q_2 - p_2)}{\sqrt{p_1q_1p_2q_2}}$$

$$\mu_4 - 3\mu_2^2 = \frac{N^4p_1p_2q_1q_2}{(N-1)^2(N-2)(N-3)} [N^2(1 - 6p_1q_1 - 6p_2q_2 + 30p_1p_2q_1q_2) + N(6p_1q_1 + 6p_2q_2 - 36p_1p_2q_1q_2) - 1]$$

$$\beta_2 - 3 = \frac{N}{(N-2)(N-3)} \left[ \left(1 - \frac{1}{N}\right) \left(\frac{1}{p_1q_1} - 6\right) \left(\frac{1}{p_2q_2} - 6\right) - 6 + \frac{1}{N} \left(1 - \frac{1}{N}\right) \frac{1}{p_1p_2q_1q_2} \right].$$



If we set

$$p = \frac{a}{a+b}, \quad p' = \frac{c}{c+d}, \quad p - p' = \frac{ad - bc}{(a+b)(c+d)} = \frac{Nx}{(a+b)(c+d)}$$

the characteristics of the distribution of  $p - p'$ , i.e., of the difference between the proportions observed in the first and second rows of the tables when the tables are restricted to the linear series used by Yates and Fisher, may be obtained at once, e.g.,

$$\mu_2(p - p') = \frac{N^2 p_1 p_2 q_1 q_2}{(N+1)(a+b)^2(c+d)^2} = \frac{1}{N-1} \frac{p_2 q_2}{p_1 q_1}.$$

If the two rows had arisen by independent sampling in two universes with probabilities  $p$  and  $p'$ , respectively, and with numbers  $n = a + b$ ,  $n' = c + d$  in the two samples, there being  $(n+1)(n'+1)$  different complexions for such pairs of samples, the value of  $\mu_2$  for the observed values of  $p - p'$  would be  $pq/n + p'q'/n'$  and if it should be assumed that  $p$  and  $p'$  were equal and had the value  $(a+c)/N$  obtained from the marginal totals we should get  $\mu_2(p - p') = p_2 q_2 / (N p_1 q_1)$  which is slightly less than that obtained above. If it were assumed that the  $2 \times 2$  table had arisen by sampling from a non-associated universe with probabilities  $p_1 p_2$ ,  $p_1 q_2$ ,  $p_2 q_1$ ,  $q_1 q_2$  for the four cells occupied by  $a$ ,  $b$ ,  $c$ ,  $d$ , respectively, the relative frequencies  $p$ ,  $p'$  in the two rows as observed could have no frequency distributions because  $N+1$  of the  $(N+3)(N+2)(N+1)/6$  samples would turn out with  $a+b=0$  and  $N+1$  of them with  $c+d=0$  and the ratios for  $p$  or  $p'$  would be  $0/0$ ; the total probability that  $p - p'$  would be indeterminate would ordinarily be small but the fact that there would be indetermination is sufficient to show that there would logically be no distribution for  $p - p'$  and hence no moments.<sup>4</sup> However, the expression  $x = (ad - bc)/N$ , though no longer interpretable in terms of the difference  $p - p'$  would have a distribution, not of  $c+d+1$  elements nor of  $(a+b+1)(c+d+1)$  but of  $(N+3)(N+2)(N+1)/6$ . The value of  $\mu_2$  for this distribution is readily found to be<sup>5</sup>

$$\mu_2(x) = (N-1)p_1 p_2 q_1 q_2 \quad \text{compared with } N p_1 p_2 q_1 q_2 \quad \text{or } \frac{N^2}{N-1} p_1 p_2 q_1 q_2$$

when the sampling is by the double point binomial or by the Yates-Fisher series, respectively, the values of the unknown probabilities in the universes in question being taken from the marginal totals.

The argument by which the linear series of tables with their relative probabilities are obtained from a given  $2 \times 2$  table by considering variations in the body of the table consistent with the given marginal totals is equally applicable to  $m \times n$  tables.<sup>6</sup> Given the table

$$\begin{array}{c|c}
 a_{11} & a_{12} & \dots & a_{1m} & A_1 \\
 a_{21} & a_{22} & \dots & a_{2m} & A_2 \\
 \dots & \dots & \dots & \dots & \dots \\
 a_{n1} & a_{n2} & \dots & a_{nm} & A_n \\
 \hline
 B_1 & B_2 & & B_m & N
 \end{array}
 \quad \text{then } \frac{\Pi_i A_i! \Pi_j B_j!}{N!} \left[ \frac{1}{\Pi_{ij} a_{ij}!} \right]$$

if  $\Pi$  be the sign of a product, is its relative probability in the  $(m-1)(n-1)$  dimensional spread<sup>7</sup> of tables derived therefrom by all variations of  $a_{ij}$  consistent with the marginal totals  $A_i$  and  $B_j$ . The mean value in such a spread for the term  $a_{ij}$  is  $A_i B_j / N$ . The average value of  $a_{ij}(a_{ij} - 1)$  or of  $a_{ij}a_{kl}$ ,  $k \neq j$ , or of  $a_{ij}a_{kl}$ ,  $k \neq i$ ,  $l \neq j$ , can be found by an argument entirely similar to that given before. For example,

$$\overline{a_{12}a_{34}} = \frac{\Pi_i A_i! \Pi_j B_j!}{N!} \left[ \sum \frac{a_{12}a_{34}}{\Pi_{ij} a_{ij}!} = \sum' \frac{1}{(a_{12} - 1)!(a_{34} - 1)! \Pi' a_{ij}!} \right]$$

where  $\sum'$  extends to all variations consistent with the former marginal totals except that  $A_1$  is replaced by  $A_1 - 1$ ,  $B_2$  by  $B_2 - 1$ ,  $A_3$  by  $A_3 - 1$ ,  $B_4$  by  $B_4 - 1$  and  $\Pi'$  extends over the  $mn - 2$  elements other than  $a_{12}$  and  $a_{34}$ .

$$\overline{a_{12}a_{34}} = \frac{A_1 B_2 A_3 B_4}{N(N-1)}$$

$$\overline{a_{12}a_{34}} - \bar{a}_{12}\bar{a}_{34} = \frac{A_1 B_2 A_3 B_4}{N(N-1)} - \frac{A_1 B_2}{N} \frac{A_3 B_4}{N}$$

or the mean product relative to the mean is

$$\pi(a_{ij}a_{kl}) = \frac{N^2}{N-1} p_i q_j p_k q_l, \quad k \neq i, l \neq j$$

where  $p_i = A_i/N$ ,  $q_j = B_j/N$ . Similarly

$$\mu_2(a_{ij}) = \frac{N^2}{N-1} p_i(1-p_i)q_j(1-q_j)$$

$$\pi(a_{ij}a_{il}) = -\frac{N^2}{N-1} p_i(1-p_i)q_j q_l.$$

Hence the correlation coefficients of two terms not in the same row or column and of two terms in the same row but not in the same column are

$$r(a_{ij}a_{kl}) = + \sqrt{\frac{p_i q_j p_k q_l}{(1-p_i)(1-q_j)(1-p_k)(1-q_l)}},$$

$$r(a_{ij}a_{il}) = - \sqrt{\frac{q_j q_l}{(1-q_j)(1-q_l)}}.$$

Higher moments could be obtained, with sufficient patience, if they were needed.

It sometimes happens that the numbers in two frequency functions are really small. For example, the distribution by age of deaths from chicken pox for the two sexes as reported in Massachusetts for 1934 were

Age	<1	1	2	3	4	5-9	10-14	15-19	20-29	30-39	40-49	50-59	Total
Male	3	1	0	0	0	1	0	0	0	1	0	1	7
Female	0	2	1	0	2	0	0	0	0	0	0	0	5
Total	3	3	1	0	2	1	0	0	0	1	0	1	12

with none for all higher age groups. This as it stands without the higher age groups is a  $2 \times 12$  table; but it seems rather clear that it has to be considered as a  $2 \times 7$  table, omitting the columns with no elements.<sup>1</sup> If so considered  $\chi^2 = 9.28$  looked up under 6 degrees of freedom gives  $P = 0.17$  which is not significant. There are 130 different tables consistent with the marginal totals. The relative probability of the given table is  $1/264$ . The 130 relative probabilities which arise have the frequency distribution

prob.	$\frac{1}{792}$	$\frac{2}{792}$	$\frac{3}{792}$	$\frac{6}{792}$	$\frac{9}{792}$	$\frac{18}{792}$
freq.	18	9	40	22	26	15

The total probability of the given table and all no more probable is  $52/264 = 0.20$  which is not much larger than the value previously obtained from  $\chi^2$ . However, the variance of the age for the males is 478 with 6 degrees of freedom and for the females is 2.4 with 4 degrees of freedom which gives a ratio of 208 and the 0.01 point is at 15.2. It is far from certain that variance analysis can be relied upon; but almost any test for the significance of the standard errors of the distributions would indicate a high degree thereof—thus indicating how different the result of different tests corresponding to different formulations of the problem may be.

<sup>1</sup> Yates, F., "Contingency Tables Involving Small Numbers and the  $\chi^2$  Test." *Suppl. Jour. Roy. Statist. Soc.*, 1, 217-235 (1934).

<sup>2</sup> The tables  $\frac{n-k}{k} \mid \frac{k}{n-k}$  and  $\frac{n-k}{k} \mid \frac{k}{n-k-1}$  where  $k < 1/2n$  have the interesting property that the total probability of the given tables and of all tables in the respective series which are no more probable than these (taking both tails together) are equal and equal to

$$\frac{2(n!)^2}{(2n)!} \left\{ 1^2 + n^2 + \left[ \frac{n(n-1)}{2!} \right]^2 + \dots + \left[ \frac{n(n-1)\dots(n-k+1)}{k!} \right]^2 \right\}$$

Thus  $\frac{4}{0} \mid \frac{0}{4}$  and  $\frac{0}{4} \mid \frac{4}{0}$  together have the same probability  $\frac{1}{35}$  as  $\frac{4}{0} \mid \frac{0}{3}$  and the tables  $\frac{5}{1} \mid \frac{1}{5}, \frac{6}{0} \mid \frac{0}{6}$  have the same probability as  $\frac{5}{1} \mid \frac{1}{4}, \frac{6}{0} \mid \frac{1}{5}, \frac{5}{5} \mid \frac{0}{0}$ , namely,  $37/402$ .

<sup>3</sup> As an illustrative example consider the table  $a = 10, b = 3, c = 2, d = 15, N = 30$ . The 13 relative probabilities are as 1; 102; 2992; 37,400; 235,620; 816,816; 1,633,632; 1,925,352; 1,337,050; 534,820; 116,688; 12,376; 476, the common denominator being 6,653,325. The series looks skew in the arithmetical sense, but actually  $\kappa = 0.02113$ ; also  $\beta_2 - 3 = -0.08084$ . If we note that for a point binomial the skewness  $\kappa$  is  $(q - p)/\sqrt{Npq}$  we note that for  $x$  in the Yates-Fisher series  $\kappa = \sqrt{N-1}(1-2/N)\kappa_1\kappa_2$  and thus varies for large values of  $N$  jointly as the square root of  $N$  and the product of the skewnesses of the point binomials in the margins.

<sup>4</sup> If  $N = 100$ , and  $p_1 = 0.02$  for the universe,  $p_2$  being anything, there would be 2 elements on the average in the first row which would, however, be entirely vacant in  $e^{-2}$  or 13.5% of the samples.

<sup>5</sup> Had we assumed a general four-fold universe the average value of  $x = (ad - bc)/N$  and of its variance would have been

$$\bar{x} = (N-1)(\pi_1 - p_1p_2), \quad \mu_2(x) = (N-1) \left[ p_1p_2q_1q_2 - \left(1 - \frac{2}{N}\right)\delta^2 + \left(1 - \frac{1}{N}\right)\delta(q_1 - p_1)(q_2 - p_2) \right]$$

where

$$\delta = \pi_1 - p_1p_2 \text{ and the universe is } \frac{\pi_1}{p_2 - \pi_1} \left| \frac{p_1 - \pi_1}{1 - p_1 - p_2 + \pi_1} \right|.$$

<sup>6</sup> Let it be assumed that  $q_1, q_2, \dots, q_m$  are the probabilities that an element fall in column 1, 2,  $\dots, m$ , respectively, in any row. Then the probability that  $a_{11}, a_{12}, \dots, a_{1m}$ , respectively, will fall in the  $n$  cells of the first row is

$$\frac{A_1! q_1^{a_{11}} q_2^{a_{12}} \dots q_m^{a_{1m}}}{a_{11}! a_{12}! \dots a_{1m}!} \text{ and } \frac{\Pi_i A_i!}{\Pi_j a_{ij}!} q_1^{B_1} q_2^{B_2} \dots q_m^{B_m}$$

will be the probability of the  $m \times n$  observed frequencies in the table. The unknown factor involving  $q_i$  will be the same for all tables having the same marginal totals. But the probability of those marginal totals arising from the distribution of  $N$  elements is

$$\frac{N!}{\Pi_j B_j!} q_1^{B_1} q_2^{B_2} \dots q_m^{B_m} \text{ and hence the quotient } \frac{\Pi_i A_i! \Pi_j B_j!}{N \Pi_{ij} a_{ij}!}$$

is the relative probability for any distribution  $a_{ij}$  consistent with the marginal totals.

<sup>7</sup> Whether "dimensions" or "degrees of freedom" or "spread" are proper terms to use for variations when numbers are very small may be questioned. One may illustrate the procedure for a particular case. Consider

$$\frac{3, 0, 2}{0, 3, 0} \text{ and its relative probability } \frac{5!(3!)^2 2!}{8!} \frac{1}{(3!)^2 2! (0!)^3} = \frac{1}{56}$$

together with all its variants and their probabilities (in brackets)

$$\begin{aligned} & \frac{3, 1, 1}{0, 2, 1} \left[ \frac{6}{56} \right]; \frac{3, 2, 0}{0, 1, 2} \left[ \frac{3}{56} \right]; \frac{2, 1, 2}{1, 2, 0} \left[ \frac{9}{56} \right]; \frac{2, 2, 1}{1, 1, 1} \left[ \frac{18}{56} \right]; \\ & \frac{2, 3, 0}{1, 0, 2} \left[ \frac{3}{56} \right]; \frac{1, 3, 1}{2, 0, 1} \left[ \frac{6}{56} \right]; \frac{1, 2, 2}{2, 1, 0} \left[ \frac{9}{56} \right]; \frac{0, 3, 2}{3, 0, 0} \left[ \frac{1}{56} \right]. \end{aligned}$$

The probability of the given table and all no more probable ones is  $2/56$ , and if one uses a level of 0.05 for significance, and uses the criterion of the given table and all no more probable ones, this table would be significant. Nothing has been said of the sampling

process by which the table arose and if I understand Yates<sup>1</sup> aright it makes no difference whether it arose by distributing 8 elements over the six cells from a universe with the probabilities  $\frac{\pi_1}{\pi_4} \left| \frac{\pi_2}{\pi_5} \right| \frac{\pi_3}{\pi_6}$  with  $\sum \pi = 1$  or by distributing 5 elements and 3 elements, respectively, to each of two trinomial universes for which  $\pi_1 + \pi_2 + \pi_3 = \pi_4 + \pi_5 + \pi_6 = 1$  or by distributing 3, 3 and 2 elements, respectively, to each of three binomial universes for which  $\pi_1 + \pi_4 = \pi_2 + \pi_5 = \pi_3 + \pi_6 = 1$ .

<sup>8</sup> If the tabulation were by sufficiently fine age groups the two rows of the contingency table would have nothing but 0's or 1's and so would the totals in the columns. If there were 7 entries of 1 in the top row and 5 entries of 1 in the bottom row and 12 entries of 1 in the marginal totals along the bottom, there would be  $12!/(7!5!) = 792$  different tables all of equal probability  $1/792$ , and the total probability of the no more probable tables would be 1.

## CONCERNING THE DIFFERENTIAL EQUATIONS OF SOME BOUNDARY LAYER PROBLEMS. II

BY HERMANN WEYL

INSTITUTE FOR ADVANCED STUDY

Communicated January 29, 1942

In a previous note under the same title<sup>1</sup> I have solved by successive approximations the following boundary value problem:

$$(A_\lambda) \begin{cases} w''' + 2ww'' + 2\lambda(k^2 - w'^2) = 0 \text{ for } z \geq 0; \\ w = w' = 0 \text{ for } z = 0, w'(\infty) = k, \end{cases}$$

involving the positive parameter  $k$ , for the special values  $\lambda = 0$  and  $1/2$ . Convergence of the process for  $\lambda = 1/2$  requires a certain constant  $q$  to be  $< 1$ . The actual computation of  $q$  is laborious; however, I wish to report that a comparatively simple estimation yields  $q < 3/4$ .

For arbitrary  $\lambda \geq 0$  I have to take refuge to a much more sophisticated method, that of fixed points of functional transformations, which was inaugurated by Birkhoff and Kellogg in a famous memoir<sup>2</sup> and later formulated in general abstract terms by Schauder and Leray. The experiences with the special cases suggest the following procedure. We eliminate the parameter  $k$  from the differential equation by differentiation,

$$w'''' + 2ww''' + 2(1 - 2\lambda)w'w'' = 0,$$

and observe that the function  $\kappa.w(\kappa z)$  solves the new equation of fourth order if  $w(z)$  does, whatever the positive constant  $\kappa$ . Hence we seek our solution of  $(A_\lambda)$  in the form  $w(z) = \kappa.f(\kappa z)$  where

$$\left. \begin{aligned} f'''' + 2ff''' + 2(1 - 2\lambda)f'f'' &= 0 \text{ for } z \geq 0; \\ f(0) = f'(0) &= 0, f''(0) = 1; f''(\infty) = 0. \end{aligned} \right\} \quad (1)$$

Our line of attack can be explained a little more easily if we replace the infinite by the finite interval  $0 \leq z \leq a$  and thus the last boundary condition in (1) by  $f''(a) = 0$ . We then introduce  $f'' = g = \varphi$  as the unknown function and start with this set-up:

$$f(z) = \int_0^z (z - \zeta)g(\zeta)d\zeta. \quad (2)$$

$$\left. \begin{aligned} \varphi'' + 2f\varphi' + 2(1 - 2\lambda)f'\varphi &= 0; \\ \varphi(0) &= 1, \varphi(a) = 0. \end{aligned} \right\} \quad (3)$$

$$\varphi = g. \quad (4)$$

For an arbitrarily given  $g$  we form (2), then solve the linear boundary value problem (3) of familiar type, thus defining the functional operator  $\Phi_\lambda$  which carries  $g$  into  $\varphi$ , and as the last step (4), we try to determine  $g$  as a function whose image coincides with the original,

$$g = \Phi_\lambda\{g\}.$$

The success hinges on the following

LEMMA: *Under the assumption  $g(z) \geq 0$  the linear problem (3) has a unique solution  $\varphi$ ; it satisfies the conditions*

$$\varphi \geq 0, \varphi' + 2f\varphi \leq 0 \quad (5)$$

throughout the interval  $0 \leq z \leq a$ .

To prove this we set

$$p(z) = \exp(2\int_0^z f(\zeta)d\zeta)$$

so that  $p(0) = 1$ ,  $p' = 2fp$ , and introduce the auxiliary function

$$\varphi_1 = p(\varphi' + 2f\varphi) = (p\varphi)'.$$

The single differential equation for  $\varphi$  is then replaced by the system

$$\left. \begin{aligned} \varphi' + 2f\varphi &= \varphi_1/p, \\ \varphi_1' - 2f\varphi_1 &= 4\lambda pf'\varphi. \end{aligned} \right\} \quad (6)$$

Multiplication by  $\varphi_1$  and  $\varphi$ , respectively, and addition leads to the formula

$$[\varphi\varphi_1]_b^a = \int_b^a \left( \frac{1}{p} \varphi_1^2 + 4\lambda pf'\varphi^2 \right) dz \quad (b < a). \quad (7)$$

The inequality  $g \geq 0$  which implies  $f' \geq 0$  guarantees the positive definite character of the "Dirichlet integral" at the right side of (7). Our lemma is an easy consequence of this fact.

Thus we have defined the operator  $\Phi_\lambda$  for all  $g \geq 0$ ; the restriction  $g \geq 0$  is harmless because it carries over to the image  $\varphi$ . Since (5) implies  $(p\varphi)' \leq 0$  we have even

$$p\varphi \leq 1, 0 \leq \varphi \leq 1,$$

and may therefore start with the set  $G$  of all functions  $g(z)$  defined and continuous in the interval  $0 \leq z \leq a$  for which  $0 \leq g \leq 1$ . This limitation of  $g$  has the consequence that

$$0 \leq f' \leq z, 0 \leq f \leq 1/2z^2 \quad (8)$$

and enables us to ascertain a universal bound for  $\varphi'$ , i.e., a number  $A$  depending on  $\lambda$  and  $a$  only such that

$$0 \leq -\varphi'(z) \leq A$$

for  $0 \leq z \leq a$  and the image  $\varphi = \Phi_\lambda\{g\}$  of every  $g \in G$ . First we determine such a universal bound for  $\varphi'(0)$ . Indeed would  $\varphi$  start off at  $z = 0$  with too steep a decline it could not help hitting the ground,  $\varphi = 0$ , long before  $z$  reaches  $a$ , considering that the coefficients of its differential equation (3) are bounded by (8). Once  $\varphi'(0)$  has been penned in that differential equation does the rest. Hence the images  $\varphi$  of all elements  $g \in G$  are equi-continuous in the strict sense that  $|z_1 - z_2| \leq \epsilon$  implies  $|\varphi(z_1) - \varphi(z_2)| \leq A\epsilon$ .

We topologize the "space" of all continuous functions of  $z$  by interpreting convergence as *uniform* convergence in the interval  $0 \leq z \leq a$ . Then  $\Phi_\lambda$  is a continuous operator in  $G$  which maps  $G$  into the subset  $G_A$  consisting of all functions  $g \in G$  for which  $|z_1 - z_2| \leq \epsilon$  implies  $|g(z_1) - g(z_2)| \leq A\epsilon$ . This suffices for the existence of a "fixed point"  $g \in G$  of the operator  $\Phi_\lambda$  (see Birkhoff and Kellogg, l.c., Theorem II).

The case of the infinite interval is not essentially more difficult. It turns out that the solution  $f$  of (1) which our method yields has the property that  $f'' \geq 0$  and  $f''' + 2ff'' \leq 0$  converge to zero with  $z \rightarrow \infty$  at least as strongly as a function of the type  $e^{-\gamma z^2}$  ( $\gamma > 0$ ). Hence

$$\int_0^\infty f''(z)dz = f'(\infty) = \beta$$

converges and we may retrace the step leading from  $(A_\lambda)$  to (1).

**THEOREM.** *For given positive  $k$  the problem  $(A_\lambda)$  has a solution  $w$  whose derivative is monotone increasing from 0 to  $k$  as  $z$  travels from 0 to infinity; the second derivative decreases monotonely from*

$$w''(0) = \alpha k^{1/2} \quad (\alpha = \beta^{-1/2})$$

*to zero, approaching zero with  $z \rightarrow \infty$  at least as strongly as a function of the type  $e^{-\gamma z^2}$ .*

<sup>1</sup> These PROCEEDINGS, 27, 578-583 (1941).

<sup>2</sup> Birkhoff, G. D., and Kellogg, O. P., *Trans. Am. Math. Soc.*, 23, 96-115 (1922); Schauder, J., *Studia Mathematica*, 2, 171-180 (1930); Leray, J., and Schauder, J., *Ann. Sc. Ec. Norm. Sup.*, 51, 45-78 (1934).

# UNSTABLE MINIMAL SURFACES WITH ANY RECTIFIABLE BOUNDARY

BY MAX SHIFFMAN

COLLEGE OF THE CITY OF NEW YORK

Communicated December 17, 1941

1. *Introduction.*—Let  $\Gamma$  be any rectifiable curve in space. We shall prove that  $\Gamma$  must bound unstable minimal surfaces whenever it bounds several minimal surfaces which are relative minima in area. More generally, the Morse relations apply in a slightly modified form. In previous literature, these results were obtained for special types of rectifiable curves.<sup>1</sup>

The method is first to obtain unstable minimal surfaces bounded by curves approximating  $\Gamma$  and second to perform a passage to the limit. The first step may be done in several ways. In the present paper it is carried through by extending the procedure used by Courant for polygonal boundaries.<sup>2</sup> The passage to the limit is accomplished by means of theorem 1 below, a theorem which was discovered by Morse and Tompkins,<sup>3</sup> and which is here proved by using the isoperimetric inequality for minimal surfaces.<sup>4</sup>

Our method has the advantage of not requiring any explicit representation of the Dirichlet integral of a potential surface in terms of its boundary values.

2. *Continuity of Area of Minimal Surfaces.*—2.1. Let  $\mathfrak{x}(r, \theta)$  be a potential surface defined over the unit circle  $r \leq 1$ , with boundary values  $\mathfrak{x}(1, \theta)$  which are continuous and of bounded variation. Indicate the length of the curve  $\mathfrak{x}(r, \theta)$ , for fixed  $r$  and  $0 \leq \theta \leq 2\pi$ , by  $L(r)$ ; indicate the length of  $\mathfrak{x}(r, \theta)$  for fixed  $\theta$  and  $\rho \leq r \leq 1$  by  $L(\theta; \rho, 1)$ . The symbol  $D[\mathfrak{x}]$  stands for the Dirichlet integral

$$D[\mathfrak{x}] = \frac{1}{2} \int \int \left( \mathfrak{x}_r^2 + \frac{1}{r^2} \mathfrak{x}_\theta^2 \right) r dr d\theta.$$

LEMMA 1:  $L(r) \leq L(1)$ .

LEMMA 2: Let  $\mathfrak{x}^n(r, \theta)$ ,  $n = 1, 2, \dots$ , be a sequence of potential surfaces converging uniformly<sup>5</sup> to the potential surface  $\mathfrak{x}^\infty(r, \theta)$ , and suppose that  $L^n(1) \rightarrow L^\infty(1)$  as  $n \rightarrow \infty$ . Then for any  $\epsilon$  there is a  $\delta$  such that any arc concentric to the unit circle whose angle is less than  $\delta$  is mapped by each  $\mathfrak{x}^n$ ,  $n = 1, 2, \dots, \infty$ , into curves of length less than  $\epsilon$ .

LEMMA 3:  $L(\theta; \rho, 1) \leq \left( \frac{4D[\mathfrak{x}]}{\delta} \cdot \log \frac{1}{\rho} \right)^{1/4}$  except for a set of  $\theta$ 's of total measure  $< \delta/2$ .

Lemma 1 is proved by use of the Poisson integral. Lemma 2 is a conse-



quence of Lemma 1 and the lower semi-continuity of length. Lemma 3 is obtained by using well-traversed estimates of  $D[\mathfrak{x}]$ .<sup>6</sup>

Define a  $\delta$ -dense radial network in the annular ring  $\rho \leq r \leq 1$  as a set of at most  $\left\lceil \frac{4\pi}{\delta} \right\rceil + 1$  radial lines  $\theta = \text{constant}$ , such that two successive radial lines have an angular difference of at most  $\delta$ . Lemma 3 shows that there is a  $\delta$ -dense radial network in the annular ring  $\rho \leq r \leq 1$  on each line of which the inequality of Lemma 3 holds.

2.2. We consider a minimal surface  $\mathfrak{x}(r, \theta)$  as given in isometric representation, i.e.,  $\mathfrak{x}$  is a potential surface and  $E = G$ ,  $F = 0$  where  $E = \mathfrak{x}_r^2$ ,  $F = \mathfrak{x}_r \cdot 1/r \mathfrak{x}_\theta$ ,  $G = 1/r^2 \mathfrak{x}_\theta^2$ . The area of a minimal surface is equal to its Dirichlet integral.

THEOREM 1.<sup>7</sup> Let  $\mathfrak{x}^n(r, \theta)$ ,  $n = 1, 2, \dots$ , be a sequence of minimal surfaces converging uniformly<sup>8</sup> to the minimal surface  $\mathfrak{x}^\infty(r, \theta)$ , and suppose that  $L^n(1) \rightarrow L^\infty(1)$  as  $n \rightarrow \infty$ . Then

$$D[\mathfrak{x}^n] \rightarrow D[\mathfrak{x}^\infty] \text{ as } n \rightarrow \infty.$$

*Proof.* Let  $M$  be an upper bound of the quantities  $L^n(1)$ ,  $[L^n(1)]^2/4\pi$ ,  $n = 1, 2, \dots, \infty$ . By the isoperimetric inequality  $A \leq L^2/4\pi$  for minimal surfaces, we have  $D[\mathfrak{x}^n] \leq M$  for all  $n$ .

For any given  $\epsilon$  choose the  $\delta$  of Lemma 2, and then  $\rho$  so near 1 that

$$\frac{2M}{\pi} \left( \frac{4M}{\delta} \cdot \log \frac{1}{\rho} \right)^{1/2} + \left( \frac{4\pi}{\delta} + 1 \right) \frac{4M}{\pi\delta} \cdot \log \frac{1}{\rho} < \epsilon.$$

Consider any  $\mathfrak{x}^n$ ,  $n = 1, 2, \dots, \infty$ . For this  $\mathfrak{x}^n$ , there is a  $\delta$ -dense radial network in the annular ring  $\rho \leq r \leq 1$  satisfying the inequality of Lemma 3. Consider the  $i$ th cell of this network. Indicate the Dirichlet integral of  $\mathfrak{x}^n$  over this cell by  $D_i'$ ; and indicate the lengths of the images under  $\mathfrak{x}^n$  of the four boundary lines of this cell by  $a_i, c_i, b_i, c_i - 1$ , where  $c_i, c_i - 1$  correspond to the radial lines,  $a_i$  to the arc of the unit circle and  $b_i$  to the arc of the circle of radius  $\rho$ . Set  $l_i = a_i + b_i + c_i + c_i - 1$ . We have  $a_i, b_i < \epsilon$  by Lemma 2 and the choice of  $\delta$ , and  $c_i, c_i - 1 < \left( \frac{4M}{\delta} \cdot \log \frac{1}{\rho} \right)^{1/2}$ . By the isoperimetric inequality,

$$\begin{aligned} 4\pi D_i' &\leq l_i^2 = (l_i + c_i + c_i - 1)(a_i + b_i) + (c_i + c_i - 1)^2 \leq \\ &\leq \left[ 2\epsilon + 4 \left( \frac{4M}{\delta} \cdot \log \frac{1}{\rho} \right)^{1/2} \right] (a_i + b_i) + \frac{16M}{\delta} \cdot \log \frac{1}{\rho}. \quad (1) \end{aligned}$$

Denote the Dirichlet integral of  $\mathfrak{x}^n$  over the annular ring  $\rho \leq r \leq 1$  by  $D'[\mathfrak{x}^n]$ ; over the circle  $r \leq \rho$  by  $D''[\mathfrak{x}^n]$ . Summing the inequality (1) for at most  $\left\lceil 4\pi/\delta \right\rceil + 1$  terms, and making use of  $\sum a_i = L^n(1) \leq M$  and  $\sum b_i = L^n(\rho) \leq L^n(1) \leq M$ , we obtain

$$D'[\xi^n] \leq \frac{M\epsilon}{\pi} + \frac{2M}{\pi} \left( \frac{4M}{\delta} \cdot \log \frac{1}{\rho} \right)^{1/2} +$$

$$\left( \frac{4\pi}{\delta} + 1 \right) \frac{4M}{\pi\delta} \log \frac{1}{\rho} < (M+1)\epsilon. \quad (2)$$

This inequality holds for  $n = 1, 2, \dots, \infty$ .

Choose  $N$  so large that  $|D''[\xi^n] - D''[\xi^\infty]| < \epsilon$  for all  $n \geq N$ . Inequality (2) yields

$$|D[\xi^n] - D[\xi^\infty]| < (2M+3)\epsilon \text{ for all } n \geq N.$$

The proof of the theorem is completed.

### 3. Unstable Minimal Surfaces Bounded by Curves Approximating $\Gamma$ .—

3.1. Let  $\mathfrak{P}$  be the space of potential surfaces  $\xi(r, \theta)$  defined over the unit circle, having a finite Dirichlet integral, mapping the circumference monotonically and continuously onto  $\Gamma$ , and satisfying the three point condition. We define the distance between two surfaces  $\xi_1, \xi_2$  as<sup>8</sup>

$$|\xi_1 - \xi_2| = \text{Max.}_{r \leq 1/2} |\xi_1(r, \theta) - \xi_2(r, \theta)|.$$

We shall first obtain unstable minimal surfaces in a space  $\bar{\mathfrak{R}}$  containing  $\mathfrak{P}$ , and then take the limit as  $\bar{\mathfrak{R}} \rightarrow \mathfrak{P}$ .

Let  $A = \{A_1, A_2, \dots, A_m\}$  be a collection of  $m$  distinct points arranged in this order on the oriented curve  $\Gamma$ . Designate the maximum among the diameters of the arcs  $A_1A_2, A_2A_3, \dots, A_mA_1$  of  $\Gamma$  by  $\sigma(A)$ . The collection  $A$  is fixed in this section.

Let  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$  be a collection of  $m$  distinct points arranged in this order on the circumference. Define the following three sets of potential surfaces:

(i)  $\mathfrak{P}(\alpha)$  is the set of all surfaces  $\xi$  of  $\mathfrak{P}$  for which  $\xi(\alpha_j) = A_j, j = 1, 2, \dots, m$ .

(ii)  $\mathfrak{R}(\alpha)$  is the set of all convex linear combinations  $\eta(r, \theta)$  of surfaces in  $\mathfrak{P}(\alpha)$ , i.e.,  $\eta$  is expressible as

$$\eta = \sum a_i \xi_i \text{ where } a_i \geq 0, \sum a_i = 1, \text{ and each } \xi_i \text{ is in } \mathfrak{P}(\alpha).$$

(iii)  $\bar{\mathfrak{R}}(\alpha)$  is the set of all surfaces  $\mathfrak{z}(r, \theta)$  with finite Dirichlet integral obtainable as limits of sequences of surfaces in  $\mathfrak{R}(\alpha)$ , i.e.,  $D[\mathfrak{z}] < \infty$  and  $\mathfrak{z} = \lim \eta^n$  where each  $\eta^n$  is in  $\mathfrak{R}(\alpha)$ . The set  $\bar{\mathfrak{R}}(\alpha)$  may be described as the *convex closure* of the set  $\mathfrak{P}(\alpha)$ .

Concerning  $\bar{\mathfrak{R}}(\alpha)$ , the following statements are easily proved.

(a)  $\mathfrak{P}(\alpha) \subset \mathfrak{R}(\alpha) \subset \bar{\mathfrak{R}}(\alpha)$ .

(b) If  $\mathfrak{z}$  is in  $\bar{\mathfrak{R}}(\alpha)$ ,  $\mathfrak{z}(1, \theta)$  is a continuous curve whose length is at most the length  $L$  of  $\Gamma$ ; and  $\mathfrak{z}(1, \theta)$  for  $\alpha_{i-1} \leq \theta \leq \alpha_i$  lies in the convex closure

of the arc  $\widehat{A_{i-1}A_i}$  of  $\Gamma$ . If  $\mathfrak{z}$  is any surface in  $\mathfrak{P}(\alpha)$ ,  $\max_{r \leq 1} |\mathfrak{z}(r, \theta) - \mathfrak{z}(r, \theta)| \leq \sigma(A)$ .

(c) The set  $\mathfrak{R}(\alpha)$  is convex, i.e., if  $\mathfrak{z}_1, \mathfrak{z}_2$  belong to  $\mathfrak{R}(\alpha)$  so does  $\mathfrak{z} = (1-t)\mathfrak{z}_2 + t\mathfrak{z}_1$  for  $0 \leq t \leq 1$ .

(d) The set  $\mathfrak{R}(\alpha)$  is  $D$ -compact, i.e., if  $\mathfrak{z}^n$  is a sequence belonging to  $\mathfrak{R}(\alpha)$  for which  $D[\mathfrak{z}^n] \leq M$ , a subsequence can be found which converges to a surface  $\mathfrak{z}$  in  $\mathfrak{R}(\alpha)$ .

(e) Let  $\mathfrak{z}$  belong to  $\mathfrak{R}(\alpha)$ , and vary  $\mathfrak{z}$  by setting

$$Z'(r, \theta) = \mathfrak{z}(r, \theta) \text{ for } \theta = \theta + \epsilon \lambda(r, \theta)$$

where  $\lambda(r, \theta)$  has continuous first derivatives and  $\epsilon$  is chosen so small that  $|\epsilon \lambda_\theta| \leq 1/2$ . Transform  $Z'$  by a linear transformation so that the three point condition is satisfied, and define  $\mathfrak{z}'$  as the potential surface having the same boundary values as the resulting surface. Set  $\alpha' =$  image of  $\alpha$  under these successive transformations. Then  $\mathfrak{z}'$  belongs to  $\mathfrak{R}(\alpha')$ .

3.2. We are ready to define the space  $\mathfrak{R}$ .  $\mathfrak{R}$  consists of all the surfaces in the  $\mathfrak{R}(\alpha)$  for all possible positions of the collection  $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ . The distance between two surfaces  $\mathfrak{z}_1, \mathfrak{z}_2$  of  $\mathfrak{R}$  is again defined by

$$|\mathfrak{z}_1 - \mathfrak{z}_2| = \max_{r \leq 1/2} |\mathfrak{z}_1(r, \theta) - \mathfrak{z}_2(r, \theta)|.$$

By property 3.1 (a),  $\mathfrak{P} \subset \mathfrak{R}$ . Property 3.1 (e) shows that it is possible to perform variations in the space  $\mathfrak{R}$ .

By using properties 3.1 (c), (d), (e) and methods similar to [3] the following theorems can be proved:

**THEOREM 2.** *The problem of minimizing  $D[\mathfrak{z}]$  among all surfaces  $\mathfrak{z}$  in  $\mathfrak{R}(\alpha)$  has a unique solution, denoted by  $\mathfrak{z}(\alpha)$ .*

**THEOREM 3.** *The surface  $\mathfrak{z}(\alpha)$  and the quantity  $d(\alpha) = D[\mathfrak{z}(\alpha)]$  depend continuously on the collection  $\alpha = \{\alpha_1, \dots, \alpha_m\}$ .*

As in [3] it may also be proved that  $d(\alpha)$  is a differentiable function of the  $\alpha_1, \alpha_2, \dots, \alpha_m$ .

Define the subspace  $\mathfrak{Q}$  of  $\mathfrak{R}$  as the set of surfaces  $\mathfrak{z}(\alpha)$  for all possible  $\alpha$ .

Theorem 2 and property 3.1 (c) assert that  $\mathfrak{R}$  may be  $D$ -retracted into  $\mathfrak{Q}$ , i.e., retracted in such a way that the Dirichlet integral never increases. Theorem 3 asserts that  $\mathfrak{Q}$  is topologically equivalent to the space of all possible  $\alpha$ 's, and that the Dirichlet integral is continuous in  $\mathfrak{Q}$ .

Thus, so far as Morse theory in  $\mathfrak{R}$  is concerned, we may limit ourselves to the space  $\mathfrak{Q}$ ; and Morse theory applies to the minimal surfaces in  $\mathfrak{Q}$ .

In particular, let  $\mathfrak{z}_1, \mathfrak{z}_2$  be any two surfaces in  $\mathfrak{R}$ . The surfaces  $\mathfrak{z}_1, \mathfrak{z}_2$  can both be  $D$ -retracted into surfaces  $\mathfrak{z}_1, \mathfrak{z}_2$  in  $\mathfrak{Q}$ . In  $\mathfrak{Q}$ ,  $\mathfrak{z}_1$  and  $\mathfrak{z}_2$  can be joined by a minimizing connected set  $C$ ; the maximum  $d$  of the Dirichlet integral on  $C$  is the smallest possible among all connected sets in  $\mathfrak{Q}$  joining  $\mathfrak{z}_1, \mathfrak{z}_2$ . If  $d > \max\{D[\mathfrak{z}_1], D[\mathfrak{z}_2]\}$ , then  $C$  contains a minimal surface  $\mathfrak{z}$  for which

$D[\xi] = d$ . Because of the isoperimetric inequality for minimal surfaces, we have  $d \leq L^2/4\pi$  by 3.1 (b).

**THEOREM 4.** *Two surfaces  $\xi_1, \xi_2$  in  $\mathfrak{R}$  can be joined by a connected set  $C$  of surfaces in  $\bar{\mathfrak{R}}$  on which the maximum of  $D[\xi]$  is a quantity  $d \leq \max. \{D[\xi_1], D[\xi_2], L^2/4\pi\}$ . If  $d > \max. \{D[\xi_1], D[\xi_2]\}$ , then  $C$  contains a minimal surface  $\xi$  for which  $D[\xi] = d$ .*

4. *Passage to the Limit to Yield Unstable Minimal Surfaces Bounded by  $\Gamma$ .*  
—4.1. The considerations in §3 depended on the selection of the collection of points  $A = \{A_1, A_2, \dots, A_m\}$  on  $\Gamma$ . The space  $\bar{\mathfrak{R}}$  defined there will now be denoted by  $\bar{\mathfrak{R}}(A)$ .

Select a sequence  $A^{(1)}, A^{(2)}, \dots, A^{(\nu)}, \dots$  of collections of points on  $\Gamma$ , where  $A^{(\nu+1)}$  is obtained from  $A^{(\nu)}$  by adding points of  $\Gamma$ , and where  $\sigma(A^{(\nu)}) \rightarrow 0$  as  $\nu \rightarrow \infty$ . Then  $\bar{\mathfrak{R}}(A^{(1)}) \supset \bar{\mathfrak{R}}(A^{(2)}) \supset \dots \supset \bar{\mathfrak{R}}(A^{(\nu)}) \supset \dots \supset \mathfrak{P}$ . And using 3.1 (b),  $\lim_{\nu \rightarrow \infty} \bar{\mathfrak{R}}(A^{(\nu)}) = \mathfrak{P}$ .

4.2. **MAIN THEOREM I.** *If the rectifiable closed Jordan curve  $\Gamma$  bounds two minimal surfaces which are proper relative minima, then  $\Gamma$  must bound an unstable minimal surface.*

*Proof.* Let the two minimal surfaces be  $\xi_1, \xi_2$ . By Theorem 4, there is a connected set  $C^{(\nu)}$  of surfaces in  $\bar{\mathfrak{R}}(A^{(\nu)})$  joining  $\xi_1, \xi_2$  on which the maximum of  $D[\xi]$  is a quantity  $d^{(\nu)}$  satisfying the inequality of Theorem 4. Select a sequence of  $A^{(\nu)}$  for which the  $d^{(\nu)}$  converge, and indicate it still by  $A^{(\nu)}$ . Consider all the limiting surfaces of the sets  $C^{(\nu)}$  as  $\nu \rightarrow \infty$ . They form a connected set  $C$  in  $\mathfrak{P}$  containing  $\xi_1, \xi_2$  on which the maximum of  $D[\xi]$  is  $d \leq \lim_{\nu \rightarrow \infty} d^{(\nu)} \leq L^2/4\pi$ .

Now,  $d > \max. \{D[\xi_1], D[\xi_2]\}$  because  $\xi_1, \xi_2$  are proper relative minima. Hence  $d^{(\nu)} > \max. \{D[\xi_1], D[\xi_2]\}$  for all  $\nu$  sufficiently large. By Theorem 4 there is a minimal surface  $\xi^*$  on  $C^{(\nu)}$  for which  $D[\xi^*] = d^{(\nu)}$ . A subsequence of the  $\xi^*$  converges to a minimal surface  $\xi$  on  $C$ . The lower semi-continuity of length and property 3.1 (b) show that the lengths of the curves  $\xi^*(1, \theta)$  converges to the length  $L$  of  $\Gamma$ . Theorem 1 yields

$$\lim_{\nu \rightarrow \infty} d^{(\nu)} = D[\xi] \leq d.$$

Hence  $d = \lim_{\nu \rightarrow \infty} d^{(\nu)}$  and  $D[\xi] = d$ . The minimal surface required by Main Theorem I is  $\xi$ .

5. *The Morse Relations.*—In the same way, by enlarging the space to  $\bar{\mathfrak{R}}(A^{(\nu)})$  and then passing to the limit  $\mathfrak{P}$ , one can prove the following:

(1) any  $k$ -cap in  $\mathfrak{P}$  with cap limit  $a$  is homologous on  $\mathfrak{P}_a$  to a  $k$ -cap which contains a minimal surface  $\xi$  for which  $D[\xi] = a$ ,<sup>9</sup>

(2) the connectivity numbers of  $\mathfrak{P}$  are  $R_0 = 1, R_1 = R_2 = \dots = R_n = \dots = 0$ .

In these two statements, we always consider Vietoris cycles lying on  $\mathfrak{P}_N$

for some  $N$ . The only difficulty in the proof of these statements, beyond the methods already used in §§1-4, is the possibility that surfaces in  $\mathfrak{P}$  may have boundary values which are constant on some arcs of the circumference.

Statement (1) yields a modified Morse theory; the usual requirement is that every  $k$ -cap contain a minimal surface  $\mathfrak{z}$  for which  $D[\mathfrak{z}] = a$ . But by assigning type numbers to *blocs* of minimal surfaces, the usual Morse relations hold.

<sup>1</sup> [1] Shiffman, "The Plateau Problem for Non-Relative Minima," *Ann. Math.*, **40**, 834-854 (1940); [2] Morse and Tompkins, "The Existence of Minimal Surfaces of General Critical Type," *Ibid.*, **40**, 443-472 (1940); [3] Courant, "Critical Points and Unstable Minimal Surfaces," these PROCEEDINGS, **27**, 51-57 (1941); [4] Morse and Tompkins, "Minimal Surfaces Not of Minimum Type by a New Mode of Approximation," *Ann. Math.*, **42**, 62-72 (1941); and "The Continuity of the Area of Harmonic Surfaces as a Function of the Boundary Representations," *Am. Jour. Math.*, **63**, 825-838 (1941).

<sup>2</sup> Cf. [3].

<sup>3</sup> Cf. [4].

<sup>4</sup> See Radò, "On the Problem of Plateau," *Ergeb. Math.*, **2**, 45-47 (1933).

<sup>5</sup> The uniformity of the convergence may be eliminated since it is a consequence of  $L^n(1) \rightarrow L^\infty(1)$ .

<sup>6</sup> Cf. Shiffman, "The Plateau Problem for Minimal Surfaces Which Are Relative Minima," *Ann. Math.*, **39**, 311-312 (1938).

<sup>7</sup> This theorem is capable of very wide generalizations.

<sup>8</sup> In  $\mathfrak{P}$  this metric is equivalent to the uniform metric.

<sup>9</sup>  $\mathfrak{P}_a$  consists of those surfaces  $\mathfrak{x}$  of  $\mathfrak{P}$  for which  $D[\mathfrak{x}] \leq a$ .

## ON HOMOGENEOUS MEASURE ALGEBRAS

BY DOROTHY MAHARAM

THE INSTITUTE FOR ADVANCED STUDY

Communicated January 26, 1942

1. The purpose of this note is to determine the structure of general measure algebras.

2. For any Boolean  $\sigma$ -algebra  $M$ , let  $\bar{M}$  be the least cardinal number which is the power of a  $\sigma$ -basis of  $M$ .  $M$  is *homogeneous*, if  $\bar{L} = \bar{M}$  for every (principal) ideal  $L \subseteq M$  different from the null ideal. A Boolean  $\sigma$ -algebra  $M = \{0 \leq a, b, c, \dots \leq e\}$  is a *measure algebra*, if there is defined on  $M$  a *measure function* (that is, a countably additive real non-negative function)  $\mu(a)$  such that (i)  $0 < \mu(e) < \infty$ , (ii)  $\mu(a) = 0$  if and only if  $a = 0$ . We also assume that there is no atomic element in  $M$ , i.e., (iii)  $a > 0$  implies the existence of a  $b \in M$  such that  $0 < b < a$ .

An example of a homogeneous measure algebra is the Boolean algebra  $P(\gamma)$  of all measurable sets (mod. null sets) of an infinite product space  $\Omega_\gamma = \prod_{\alpha < \gamma} I_\alpha$  of intervals  $I_\alpha: 0 \leq x_\alpha \leq 1$ , where  $\alpha$  and  $\gamma$  are ordinal numbers and the measure on  $\Omega_\gamma$  is defined multiplicatively in terms of the ordinary Lebesgue measure on each  $I_\alpha$ .

If  $\gamma < \gamma'$ , then  $P(\gamma)$  may be considered as a  $\sigma$ -subalgebra of  $P(\gamma')$ . This embedding is obtained by identifying each subset  $E$  of  $\Omega_\gamma$  with the cylinder set  $E \times \prod_{\gamma \leq \alpha < \gamma'} I_\alpha$  in the product space  $\Omega_{\gamma'} = \Omega_\gamma \times \prod_{\gamma \leq \alpha < \gamma'} I_\alpha$ .

Two measure algebras are *isomorphic*, if there exists a measure-preserving  $\sigma$ -isomorphism between them. It is easy to see that, for any infinite ordinals  $\gamma$  and  $\gamma'$ , two measure algebras  $P(\gamma)$  and  $P(\gamma')$  are isomorphic if and only if  $\gamma$  and  $\gamma'$  correspond to the same cardinal. Moreover, if  $\gamma$  is finite or corresponds to  $\aleph_0$ , then  $P(\gamma)$  is isomorphic to  $P(1)$ , i.e., to the measure algebra of Lebesgue measurable sets (mod. null sets) of the interval  $I: 0 \leq x \leq 1$ .

**THEOREM 1.** *Every homogeneous measure algebra with  $\mu(e) = 1$  is isomorphic to  $P(\gamma_0)$ , where  $\gamma_0$  is the least ordinal number corresponding to  $M$ .*

This theorem will be proved by transfinite induction. It will be sufficient to prove the following

**LEMMA 1:** *Let  $L$  be a  $\sigma$ -subalgebra of a homogeneous measure algebra  $M$  with  $\mu(e) = 1$  such that  $\bar{L} < \bar{M}$ , and assume that  $L$  is isomorphic to  $P(\gamma)$ , where  $\gamma$  is an ordinal corresponding to  $\bar{L}$ . Then, for any  $a \in M$  with  $a \notin L$ , there exists a  $\sigma$ -subalgebra  $L'$  of  $M$  such that  $L \subset L'$ ,  $a \in L'$ , and  $L'$  is isomorphic to  $P(\gamma + 1)$ , in such a way that this isomorphism is an extension of the given isomorphism of  $L$  and  $P(\gamma)$ .*

The proof of this lemma will be given in section 4.

3. Let us assume the conditions of Lemma 1. We can consider  $\Omega_\gamma$  as a representation space of  $L = P(\gamma)$ . Then, by a theorem of Radon-Nikodym, there exists, for any  $a \in M$ , a measurable function  $\varphi_a(\xi)$  defined on  $\Omega_\gamma$  such that  $0 \leq \varphi_a(\xi) \leq 1$ , and  $\int_{E_x} \varphi_a(\xi) d\xi = \mu(a \wedge x)$  for any  $x \in L$ , where  $E_x$  is a measurable set of  $\Omega_\gamma$  which corresponds to  $x \in L$ . An element  $a \in M$  is called *independent* of  $L$ , if we have  $\varphi_a(\xi) = \text{constant}$  [namely,  $= \mu(a)$ ] almost everywhere (a. e.) on  $\Omega_\gamma$ . This is equivalent to saying that we have  $\mu(a \wedge x) = \mu(a) \cdot \mu(x)$  for all  $x \in L$ .

**LEMMA 2:** *For any measurable function  $\chi(\xi)$  defined on  $\Omega_\gamma$  with  $0 \leq \chi(\xi) \leq \varphi_a(\xi)$  a. e. on  $\Omega_\gamma$ , there exists a  $b \in M$  such that  $0 \leq b \leq a$  and  $\varphi_b(\xi) = \chi(\xi)$  a. e. on  $\Omega_\gamma$ .*

*Proof of Lemma 2.*—Our lemma is clear if  $\chi(\xi) = 0$  a. e. on  $\Omega_\gamma$ . Hence we may assume that  $\text{meas}\{\xi: \chi(\xi) > 0\} > 0$ . Because of the principle of exhaustion, it is sufficient to show that there exists a  $b^* \in M$  such that  $0 < b^* \leq a$  and  $\varphi_{b^*}(\xi) \leq \chi(\xi)$  a. e. on  $\Omega_\gamma$ .

Let  $A$  be the principal ideal generated by  $a$  and let  $L(a)$  be the  $\sigma$ -subalgebra of  $M$  generated by  $L$  and  $a$ . Since  $\overline{L(a)} = \bar{L} < \bar{A} = \bar{M}$  by assumption,

there exists a  $b_1 \in M$  such that  $b_1 \in A$  and  $b_1 \bar{\in} L(a)$ . This means that  $\text{meas}[\xi: 0 < \varphi_{b_1}(\xi) < \varphi_a(\xi)] > 0$ . Again, by the principle of exhaustion, we can further find a  $b_2 \in M$  such that  $0 < b_2 < a$  and  $0 < \varphi_{b_2}(\xi) < \varphi_a(\xi)$  a. e. on the set  $[\xi: \varphi_a(\xi) > 0]$ . Let us denote by  $c_1$  and  $c_2$  the elements of  $L$  which correspond to the sets  $[\xi: 0 < \varphi_{b_1}(\xi) \leq 2^{-1}\varphi_a(\xi)]$  and  $[\xi: 2^{-1}\varphi_a(\xi) < \varphi_{b_1}(\xi) < \varphi_a(\xi)]$ . If we now put  $b_3 = (c_1 \wedge b_2) \vee (c_2 \wedge (a - b_2))$ , then  $0 < b_3 < a$  and we have  $0 < \varphi_{b_3}(\xi) \leq 2^{-1}\varphi_a(\xi)$  a. e. on the set  $[\xi: \varphi_a(\xi) > 0]$ . By iterating this argument, we can find, for each  $n$ , a  $b_{n+2} \in M$  such that  $0 < b_{n+2} < a$  and  $0 < \varphi_{b_{n+2}}(\xi) \leq 2^{-n}\varphi_a(\xi)$  a. e. on the set  $[\xi: \varphi_a(\xi) > 0]$ . Take  $n$  so large that  $\text{meas } E^* > 0$ , where  $E^* = [\xi: 2^{-n}\varphi_a(\xi) \leq \chi(\xi)]$ , and denote by  $c^*$  the element of  $L$  which corresponds to  $E^*$ . If we put  $b^* = c^* \wedge b_{n+2}$ , then we have  $0 < b^* < a$ , and  $\varphi_{b^*}(\xi) = \varphi_{b_{n+2}}(\xi) \leq 2^{-n}\varphi_a(\xi) \leq \chi(\xi)$  a. e. on  $E^*$ , and  $\varphi_{b^*}(\xi) = 0$  a. e. on  $\Omega_\gamma - E^*$ . Hence  $\varphi_{b^*}(\xi) \leq \chi(\xi)$  a. e. on  $\Omega_\gamma$  and this proves Lemma 2.

4. *Proof of Lemma 1.* Let  $\Delta_n$  be the set of all finite sequences  $\delta = \{\epsilon_1, \dots, \epsilon_n\}$ , where  $\epsilon_i = 0$  or  $1$ ,  $i = 1, \dots, n$ ; and let  $\Delta = \bigcup_{n=1}^{\infty} \Delta_n$ . A countable set  $\{a_\delta\}$  ( $\delta \in \Delta$ ) is a *dyadic decomposition* of  $a$  if  $a_0 \vee a_1 = a$ ,  $a_0 \wedge a_1 = 0$ , and  $a_{\delta,0} \vee a_{\delta,1} = a_\delta$ ,  $a_{\delta,0} \wedge a_{\delta,1} = 0$  for all  $\delta \in \Delta$ . By Lemma 2, there exists for any  $a \in M$ , a dyadic decomposition  $\{a_\delta\}$  ( $\delta \in \Delta$ ) of  $a$  such that  $\varphi_{a_\delta}(\xi) = \min.(\varphi_a(\xi), \epsilon_1/2 + \dots + \epsilon_n/2^n + 1/2^n) = \min.(\varphi_a(\xi), \epsilon_1/2 + \dots + \epsilon_n/2^n)$  a. e. on  $\Omega_\gamma$  for all  $\delta = \{\epsilon_1, \dots, \epsilon_n\} \in \Delta$ . In the same way, there exists a dyadic decomposition  $\{a'_\delta\}$  ( $\delta \in \Delta$ ) of  $a' = e - a$  such that  $\varphi_{a'_\delta}(\xi) = \max.(\varphi_a(\xi), \epsilon_1/2 + \dots + \epsilon_n/2^n + 1/2^n) = \max.(\varphi_a(\xi), \epsilon_1/2 + \dots + \epsilon_n/2^n)$  a. e. on  $\Omega_\delta$  for all  $\delta = \{\epsilon_1, \dots, \epsilon_n\} \in \Delta$ . Let us put  $b_\delta = a_\delta \vee a'_\delta$  for all  $\delta \in \Delta$ . Then  $\{b_\delta\}$  ( $\delta \in \Delta$ ) is a dyadic decomposition of the unit element  $e$ , and each  $b_\delta$  is independent of  $L$ , since we have clearly  $\varphi_{b_\delta}(\xi) = \varphi_{a_\delta}(\xi) + \varphi_{a'_\delta}(\xi) = 1/2^n$  a. e. on  $\Omega$  for all  $\delta \in \Delta_n$ ,  $n = 1, 2, \dots$ .

Now let  $L'$  be the  $\sigma$ -subalgebra of  $M$  generated by  $L$  and  $\{b_\delta\}$  ( $\delta \in \Delta$ ).  $L'$  is obtained by completing the finite algebra  $L^*$  consisting of all elements of  $M$  of the form:  $b^{(n)} = \bigvee_{\delta \in \Delta_n} (b_\delta \wedge c_\delta)$ , where  $c_\delta \in L$  for all  $\delta \in \Delta_n$ ,  $n = 1, 2, \dots$ . It is easy to see that  $L'$  is isomorphic to  $P(\gamma + 1)$  by an isomorphism which is an extension of the given isomorphism of  $L$  and  $P(\gamma)$ .

Finally, in order to prove that  $a \in L'$ , consider the set  $[\xi: \varphi_a(\xi) \geq \epsilon_1/2 + \dots + \epsilon_n/2^n + 1/2^n]$  for each  $\delta = \{\epsilon_1, \dots, \epsilon_n\} \in \Delta$ , and let  $c_\delta$  be the corresponding element of  $L$ . If we put  $b^{(n)} = \bigvee_{\delta \in \Delta_n} (b_\delta \wedge c_\delta)$ , then we have  $b^{(1)} \leq b^{(2)} \leq \dots \leq b^{(n)} \leq \dots \leq a$  and  $\mu(a - b^{(n)}) \leq 1/2^n$  for  $n = 1, 2, \dots$ . Hence we have  $a = \bigvee_{n=1}^{\infty} b^{(n)}$  and consequently  $a \in L'$ . This proves Lemma 1 and so Theorem 1.

5. A measure algebra  $M$  is a *direct sum* of (a finite or countably infinite number of) measure algebras  $M_n$ , if each  $M_n$  is (isomorphic to) a principal ideal of  $M$  and if every element  $a \in M$  is uniquely expressed in the form:



$a = \bigvee_{n=1}^{\infty} a_n$ ,  $a_m \wedge a_n = 0$  ( $m \neq n$ ), where  $a_n \in M_n$ ,  $n = 1, 2, \dots$ . (This last condition is equivalent to saying that the unit elements  $e_n$  of  $M_n$ , which are elements of  $M$ , satisfy  $e_m \wedge e_n = 0$  ( $m \neq n$ ) and  $e = \bigvee_{n=1}^{\infty} e_n$ .)

**THEOREM 2.** *Every measure algebra is a direct sum of homogeneous measure algebras  $M_n$  ( $n = 1, 2, \dots$ , finite or countably infinite).*

The proof of Theorem 2 is easy and will be omitted.

Theorems 1 and 2 indicate the structure of a general measure algebra.

6. The ergodicity of a measure preserving  $\sigma$ -automorphism  $T$  of a measure algebra  $M$  (onto itself), or that of a group  $G = \{T\}$  of such  $\sigma$ -automorphisms, can be defined as usual.

**LEMMA 3.** *In order that the group of all measure preserving  $\sigma$ -automorphisms of a measure algebra  $M$  be ergodic on  $M$ , it is necessary and sufficient that  $M$  be homogeneous.*

The proof of this lemma is easy and is omitted.

**THEOREM 3.** *Let  $M$  be a measure algebra, and let  $G$  be the group of all measure preserving  $\sigma$ -automorphisms of  $M$ . Then  $M$  is a direct sum of a countable number of invariant principal ideals  $M_n$ , on each of which  $G$  is ergodic. This decomposition is the same as in Theorem 2.*

Let  $G$  be an arbitrary group of measure preserving  $\sigma$ -automorphisms of  $M$ . Then the set  $L_G$  of all  $a \in M$  such that  $T(a) = a$  for all  $T \in G$ , is a  $\sigma$ -subalgebra of  $M$ . Conversely, for any  $\sigma$ -subalgebra  $L \subseteq M$ , consider the set  $G_L$  of all measure preserving  $\sigma$ -automorphisms  $T$  of  $M$  such that  $T(a) = a$  for all  $a \in L$ .  $G_L$  is clearly a group. It is also clear that  $L \subseteq L_{G_L}$  and  $G \subseteq G_{L_G}$ . Exactly as in the theory of Galois, we may ask the question: Under what conditions do the equalities  $L = L_{G_L}$  and  $G = G_{L_G}$  hold? These and related problems will be discussed elsewhere.

I am greatly indebted to Shizuo Kakutani, who revised and improved my original manuscript.



## ON THE DEPOLARIZATION OF NEUTRON BEAMS BY MAGNETIC FIELDS

BY O. HALPERN AND T. HOLSTEIN

RADIATION LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY AND WESTING-  
HOUSE RESEARCH LABORATORY

Communicated January 10, 1941

In two previous notes<sup>1</sup> referred to henceforth as I and II the authors have discussed the transmission of neutrons through magnetized and unmagnetized ferromagnets and have in detail reported on the physical consequences derivable from observations on polarized neutron beams. For all the applications reference should be made to II.

The mathematical method used was semi-classical; the positional coordinates  $\mathbf{r}$  of the neutron were treated fully classically because the wavelength is very small compared to the linear dimensions of the domains in which the magnetic field changes; the behavior of the spin was described with the aid of the self-explanatory equation

$$\frac{d\mathbf{s}}{dt} = g[\mathbf{s} \times \mathbf{B}(\mathbf{r})] \quad (1)$$

which could be integrated for all physically interesting cases as far as the dependence of  $\mathbf{B}$  on  $\mathbf{r}$  was concerned; the result of this integration, when averaged over all neutron paths, gave directly the change in the *average spin*.

There exist problems of considerable physical interest in which it is necessary to know not only the average spin but the average number of neutrons in each individual spin state. For this type of problem the method described and applied below seems to be appropriate. We shall limit the discussion to two cases which are (cf. II) of direct physical significance. The method is general and applies to all kinds of particles and all values of the spin.

The change in the number of neutrons occupying each spin state is determined as follows: Every individual particle undergoes changes in its spin state under the influence of the magnetic fields. The final result is obtained by averaging over all paths of the various particles which in general will traverse regions of different magnetic field. It will be seen below that the effect of the transmission is to *equalize* the average number of neutrons in each spin state; since a polarized beam is characterized by a preferential occupation of the different spin states, the effect of a ferromagnetic medium is *depolarization* of such a beam.

(1) The magnetic field  $\mathbf{B}_i$  of the  $i$ th domain which is traversed by a neutron during the time interval  $\tau_i$  is random in direction but constant in

magnitude and direction during  $\tau_i$ . The times  $\tau_i$  are in addition limited by the condition  $gB_i\tau_i \ll 1$ . This case constitutes an excellent approximation to the behavior of an unmagnetized polycrystalline specimen, whose crystal grains have linear dimensions  $\ll 10^{-8}$  cm.

The starting point is the wave equation

$$i\hbar \frac{dc_n}{dt} = \sum_m H_{nm} c_m. \quad (2)$$

Here  $c_m$  is the amplitude of the state for which  $s_z = n\hbar$  and the  $H_{nm}$  are the matrix elements of the Hamiltonian  $g\mathbf{s} \cdot \mathbf{B}$  viz.:

$$H_{nn} = n\hbar g B_z \quad (3a)$$

$$H_{n, n \pm 1} = \frac{s}{2} g\hbar (B_x \mp iB_y) \sqrt{(l \mp n)(l \pm n + 1)}. \quad (3b)$$

where

$$l = \frac{|s_z|_{\max}}{\hbar}.$$

As before, the positional coördinates are treated classically:

$$x = x_0 + vt, \quad y = y_0, \quad z = z_0.$$

The existence of randomly oriented fields is taken into account in the following manner: The total time  $T$  is divided into intervals  $\tau_i$  each of which is characterized by a Hamiltonian matrix  $H_{nm}^{(i)}$  the average of which over all neutron paths has the following properties:

$$\overline{H_{nm}^{(i)}} = 0, \quad \overline{H_{nm}^{(i)} H_{mp}^{(j)}} = \delta_{ij} \delta_{np}. \quad (4)$$

The condition  $gB_i\tau_i \ll 1$  becomes  $\frac{H_{nm}^i \tau_i}{\hbar} \ll 1$ .

One could now find exact solutions of (2), and average by (4); this procedure is somewhat cumbersome and can be avoided by the following method. One introduces a time interval  $\vartheta$  intermediate in magnitude between the  $\tau_i$  and  $T$  such that: (a) many domains are traversed in  $\vartheta$ ; (b) the relative change suffered by  $c_n$  during  $\vartheta$  is small. Using the second property, one can immediately write down the values of the  $c_n$  at a time  $t + \vartheta$  by a method of successive approximations:

$$c_n(t + \vartheta) = c_n^0(t + \vartheta) + c_n^{(1)}(t + \vartheta) + c_n^{(2)}(t + \vartheta) \quad (5)$$

where

$$c_n^0(t + \vartheta) = c_n(t) \quad (5a)$$

$$c_n^{(1)}(t + \vartheta) = \frac{1}{i\hbar} \sum_i \sum_m H_{nm}^{(i)} c_m(t) \tau_i \quad (5b)$$

$$c_n^{(2)}(t + \vartheta) = -\frac{1}{\hbar^2} \sum_i \sum_{j < i} \sum_m \sum_p H_{nm}^{(i)} H_{mp}^{(j)} c_p(t) \tau_i \tau_j - \frac{2}{\hbar^2} \sum_i \sum_m \sum_p H_{nm}^{(i)} H_{mp}^{(i)} c_p(t) \tau_i^2. \quad (5c)$$

$\sum_i$  here indicates summation over all intervals  $\tau_i$  which are enclosed in  $\vartheta$ .

The first term of (5c) represents the effect in domains ( $i$ ) of the changes in  $c_n$  produced in previously traversed domains; the second term gives the second order effect of the  $i$ th domain, itself. For the probability  $w_n(t + \vartheta) = c_n c_n^*$  one obtains, to the second order:

$$w_n(t + \vartheta) = w_n(t) + [c_n^{0*}(t) c_n^{(1)}(t + \vartheta) + \text{comp. conj.}] - \frac{1}{\hbar^2} \sum_{i, i'} \sum_{m, m'} H_{nm}^{(i)} H_{nm'}^{(i')} c_m(t) c_{m'}^*(t) \tau_i \tau_{i'} - \frac{1}{\hbar^2} \left[ \sum_i \sum_{m, p} H_{nm}^{(i)} H_{mp}^{(i)} c_p(t) c_n^*(t) + \text{compd. conj.} \right] - \frac{1}{2\hbar^2} \left[ \sum_{i, j < i} \sum_{m, p} H_{nm}^{(i)} H_{mp}^{(j)} c_p(t) c_n^*(t) \tau_i \tau_j + \text{compl. conj.} \right]. \quad (6)$$

For the average over all neutron paths, i.e., for the average over all directions of the magnetic field, one obtains by (4)

$$w_n(t + \vartheta) = w_n(t) + \frac{1}{\hbar^2} \sum_{i, m, m'} H_{nm}^{(i)} H_{nm'}^{(i)} c_m(t) c_{m'}^*(t) \tau_i^2 - \frac{1}{2\hbar^2} \left[ \sum_{i, j} \sum_{m, p} H_{nm}^{(i)} H_{mp}^{(j)} c_p(t) c_n^*(t) \tau_i \tau_j + \text{compl. conj.} \right] = w_n(t) + \frac{1}{\hbar^2} \sum_{i, m} \overline{|H_{nm}^{(i)}|^2} w_m(t) \tau_i^2 - \frac{1}{\hbar^2} \sum_{i, m} \overline{|H_{nm}^{(i)}|^2} w_n(t) \tau_i^2. \quad (7)$$

In obtaining the last term, the Hermitian property  $H_{nm} = H_{mn}^*$  was employed.

Since  $\vartheta$  is so chosen that only small changes occur in  $c_n$  and, consequently in  $w_n$ , one can replace the differences by differentials, obtaining:

$$\frac{\partial w_n}{\partial t} = \frac{1}{\hbar^2} \frac{\tau^2}{\tau_i} \sum_m \overline{|H_{nm}|^2} (w_m - w_n) \approx \frac{\tau}{\hbar^2} \sum_m \overline{|H_{nm}|^2} (w_m - w_n). \quad (8)$$

This equation is an example of a type commonly encountered in statistics. The change in the probability of occupation of a state,  $n$ , is equal to the probability of transitions from all other states,  $m$ , minus the probability of reverse transitions—from  $n$  to  $m$ . A stationary condition is reached when all the states are equally occupied. The manner in which equilibrium is attained is somewhat complicated for the general case, but can be illustrated very simply for the particular case of spin  $1/2$ . Equation (3) together with the equality of  $\overline{B_x^2}$ ,  $\overline{B_y^2}$ ,  $\overline{B_z^2}$  leads to the equations:

$$\frac{\partial w_{1/2}}{\partial t} = \frac{1}{6} g^2 \tau \overline{B^2} (w_{-1/2} - w_{1/2}) = - \frac{\partial w_{-1/2}}{\partial t}. \quad (9)$$

The connection with the classical method can be established by calculating the expectation value for the polarization. In the simple case of spin  $1/2$  this polarization  $s_z$  is  $s/2(w_{1/2} - w_{-1/2})$ ; the equation for its variation is easily obtained from (9):

$$\frac{\partial s_z}{\partial t} = - \frac{1}{3} g^2 \tau \overline{B^2} s_z \quad (10)$$

which has the solution

$$s_z = s_{z0} \exp \left( - \frac{1}{3} g^2 \tau \overline{B^2} t \right) \quad (10a)$$

to be compared with II.

The equation for the spin can readily be obtained in the general case of spin  $l$ . Through use of (3) equation (7) becomes

$$\frac{\partial w_n}{\partial t} = \frac{1}{4} \tau g^2 (\overline{B_x^2} + \overline{B_y^2}) [(l-n)(l+n+1)(w_{n+1} - w_n) + (l+n)(l-n+1)(w_{n-1} - w_n)]. \quad (11)$$

The equation for the  $z$  component of the spin is:

$$\begin{aligned} \frac{\partial s_z}{\partial t} = \frac{\partial}{\partial t} \sum_n n w_n = \frac{1}{4} g^2 \tau (\overline{B_x^2} + \overline{B_y^2}) \sum_n [n w_{n+1} (l-n)(l+n+1) + \\ n w_{n-1} (l+n)(l-n+1) - n w_n \{ (l-n)(l+n+1) + \\ (l+n)(l-n+1) \}]. \quad (12) \end{aligned}$$

Now

$$\sum_n n w_{n+1} (l-1)(l+n+1) = \sum_{n'=-l+1}^{n'=l+1} (n'-1)(l-n'+1) w_{n'} = \sum_{n=-l}^{n=l} (n-1)(l-n+1)(l+n) w_n$$

since the terms not common to both summations actually have the value zero, as can be learned from inspection. Similarly

$$\sum_{n=-l}^{n=l} n w_{n-1} (l+n)(l-n+1) = \sum_{n=-l}^{n=l} (n+1)(l-n)(l+n+1) w_n.$$

Substituting in (12) one obtains

$$\frac{\partial s_z}{\partial t} = -\frac{1}{2} \tau g^2 (\overline{B_x^2} + \overline{B_y^2}) \sum n w_n = -\frac{1}{3} \tau g^2 \overline{B^2} s_z \quad (13)$$

which has the solution

$$s_z(t) = s_{z0} \exp \left( -\frac{1}{3} \tau g^2 \overline{B^2} t \right) \quad (13a)$$

to be compared with (10a).

This formula constitutes the quantum-mechanical verification of the important fact that *the depolarization depends, not on the magnetic moment or the spin separately, but only on their ratio g*.

(2). The total field  $\mathbf{B}$  of each domain is the sum of two fields  $\mathbf{B}_0 + \mathbf{B}_i$ . The first,  $\mathbf{B}_0$ , is constant throughout the time  $T$ . The second  $\mathbf{B}_i$  is constant within  $\tau_i$ . The direction of  $\mathbf{B}_i$  is perpendicular to that of  $\mathbf{B}_0$  but otherwise random while the absolute value of  $B_i$  is small compared to that of  $B_0$ . It will furthermore be assumed that the initial direction of the spin is parallel or anti-parallel to  $\mathbf{B}_0$ . These assumptions (cf. II) approximate the conditions prevailing during the passage of neutrons through a *quasi saturated polycrystal*.

These conditions take on the following mathematical form:

$$B_{xi} = 0, \overline{B_{xi} B_{yj}} = 0 \text{ unless } i = j, x = y$$

$$\overline{B_{xi}} = \overline{B_{yi}} = 0; \overline{B_{xi}^2} = \overline{B_{yi}^2} = \frac{1}{2} \overline{B_i^2}.$$

Due to the fact that  $gB\tau_i$  is not small, an approach analogous to that of the preceding section is not applicable. In order to solve (2) explicitly, one sets up a time-dependent perturbation treatment. Introducing the quantities  $a_n = c_n \exp(-iE_n t/\hbar)$  where  $E_n = H_{nn}$  one obtains the well-known equations

$$i\hbar \frac{da_n}{dt} = \sum_m H_{nm} a_m \exp[-i(E_m - E_n)t/\hbar]. \quad (14)$$

If the axis of quantization is chosen parallel to the vector  $\mathbf{B}_0$

$$H_{nn} = E_n = n g \hbar B_0 \quad (15a)$$

$$H_{n, n \pm 1}^{(i)} = 1/2 \hbar g (B_{xi} \mp i B_{yi}) \sqrt{(l \mp n)(l \pm n + 1)} \quad (15b)$$

from which it is apparent that the usual mechanism of the time dependent perturbation theory is an expansion in the parameter  $|B_i|/|B_0|$ .

The solution of (14) is similarly to (5):

$$a_n(t + \vartheta) = a_n(t) + a_n^{(1)}(t + \vartheta) + a_n^{(2)}(t + \vartheta) \quad (16)$$

$$a_n^{(1)}(t + \vartheta) = \sum_i \sum_m H_{nm}^{(i)} a_m(t) \exp(-i(E_m - E_n)\vartheta_i/\hbar) \frac{\exp(-i(E_n - E_m)\tau_i/\hbar) - 1}{E_n - E_m} \quad (16a)$$

$$a_n^{(2)}(t + \vartheta) = \sum_i \sum_{m, p} H_{nm}^{(i)} H_{mp}^{(i)} a_p(t) \frac{\exp(-i(E_p - E_n)\vartheta/\hbar)}{E_p - E_n} \left[ \frac{\exp(-i(E_p - E_n)\tau_i/\hbar) - 1}{E_p - E_n} - \frac{\exp(-i(E_m - E_n)\tau_i/\hbar) - 1}{E_m - E_n} \right] + \sum_{i, j < i} \sum_{m, p} H_{nm}^{(i)} H_{mp}^{(j)} a_p(t) \exp[-i/\hbar \{ (E_m - E_n)\vartheta_i - (E_p - E_n)\vartheta_j \}] \frac{\exp(-i/\hbar(E_n - E_m)\tau_i) - 1}{E_n - E_m} \cdot \frac{\exp(-i/\hbar(E_m - E_p)\tau_j) - 1}{E_m - E_p} \quad (16b)$$

$$\text{with } \vartheta_i = \sum_{j=1}^{i-1} \tau_j.$$

One now sets up the probability,  $w_n(t)$ , as before, and averages over the direction of the magnetic fields  $\mathbf{B}_i$ , obtaining

$$w_n(t + \vartheta) = w_n(t) + 4 \sum_{i, m} \frac{|H_{nm}|^2}{(E_n - E_m)^2} (w_m - w_n) \sin^2(E_n - E_m)\tau_i/2\hbar. \quad (17a)$$

Replacement of finite differences by differentials yields, finally:

$$\frac{dw_n}{dt} = \sum_m \frac{|H_{nm}|^2}{(E_n - E_m)^2} \frac{4}{\tau} (w_m - w_n) \sin^2(E_n - E_m)\tau_i/2\hbar. \quad (17b)$$

Equation (17b) can be integrated without particular difficulty for any case of actual interest. One can on the other hand learn from it without

calculation how the two qualitatively different results of II were arrived at. We found that for  $gB\tau \ll 1$  the depolarization coefficient is proportional to  $B_i^2$  independent of  $B_0$  while in the other case it turned out to be proportional to  $B_i^2/B_0^2$ . These results are here immediately obtained from (17b) since it is legitimate in the first case to expand the  $\sin^2$  thereby removing the energy denominator  $(E_n - E_m)^2$  which is proportional to  $B_0^2$ . In the second case  $\sin^2$  may be replaced by  $1/2$  leaving thereby the energy denominator  $\sim B_0^{-2}$ .

These mathematical features permit an informative physical interpretation. In each time interval  $\tau$ , matrix elements  $\sim gB_i\hbar$  cause transitions between energy levels which are separated by  $\sim g\hbar B_0$ . The uncertainty principle  $\Delta E\tau \sim \hbar$  when applied to this case states that transitions occur with appreciable frequency only between such states which permit conservation to within  $\hbar/\tau$ . If therefore  $gB_0\hbar > \hbar/\tau$  then transitions are cut down; which leads mathematically to a polarization coefficient proportional to  $B_i^2/B_0^2$ . If, however,  $gB_0\hbar \ll \hbar/\tau$ , then no restrictions on the frequency of transitions are imposed by the energy principle since the quantum theoretical uncertainty exceeds the energy differences between the various states. The frequency of transitions is then solely determined by the values of the non-diagonal matrix elements which are proportional to  $B_i$  and independent of  $B_0$ .

<sup>1</sup> Halpern, O., and Holstein, T., (I) *Phys. Rev.*, **55**, 601 (1939); (II) *Ibid.*, **59**, 960 (1941).

# PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

Volume 28

April 15, 1942

Number 4

*Copyright 1942 by the National Academy of Sciences*

---

## *AN AUTOSOMAL RECESSIVE FACTOR INDUCING SEMI-STERILITY IN DROSOPHILA MELANOGASTER FEMALES*

BY U. FANO

CARNEGIE INSTITUTION OF WASHINGTON, DEPARTMENT OF GENETICS,  
COLD SPRING HARBOR, N. Y.

Communicated March 14, 1942

Measurements of the frequency of "dominant lethals" induced in sperm of *Drosophila*<sup>1</sup> have been seriously hampered by appreciable uncontrolled discrepancies between the results of parallel experiments. It was found that when untreated material was used in mass matings, the number of eggs that failed to develop into adult flies showed non-random variations from zero up to about 40%, averaging about 20%. The failures in development such as were observed in eggs from untreated parents are consequently superimposed on the induced failures to be measured in experiments on "dominant lethals." Reducing the non-random failures in development of untreated material to a negligible amount should, therefore, greatly improve the accuracy of experiments on "dominant lethals." With this in mind, an effort was made to investigate the causes for failures in development of untreated material. The results are reported in this paper.

*Experimental Methods.*—The methods used varied somewhat in different experiments. Eggs were collected from single females. Failures of development were observed by counting either the relative numbers of hatched and unhatched eggs or the number of adult flies emerging per 100 eggs collected. In the former case the most satisfactory procedure was to collect eggs which had been deposited on corn meal agar (coated with fresh yeast) spread on a wooden spoon. The spoons were then stored for incubation in empty vials, proper care being taken to ensure sufficient moisture. In the latter case it proved advisable to transfer the eggs to a corn meal agar slant within a vial, in order to minimize the chances of accidental deaths. When only the presence or absence of a high sterility effect had to be determined (see further below), it was sufficient to collect eggs on corn meal agar within a vial and to examine the vial after incubation to see whether most of the eggs had or had not hatched.



*Results.*—Preliminary observation of the failures in development of untreated material showed the following points: (a) The phenomenon can already be observed at the egg-hatching stage. (b) A large share of the failures in development occurs among the offspring of comparatively few females ("semi-sterile" females); the non-random occurrence of unhatched eggs in mass cultures is thus related to the variable number of eggs laid by "semi-sterile" females within any one experimental sample. (c) The frequency of failures in development among the offspring of "semi-sterile" females ranges anywhere up to 100%; successive tests on eggs laid by the same female yield comparable results, indicating that the failures are due to biological causes and not to experimental accidents. (d) "Semi-sterile" females do not seem to be confined to the inbred wild type *Swedish b* stock on which most of the experiments were made. (e) "Semi-sterility" is not affected when a female is outbred to a male from a different stock, and therefore it appears to be an actual characteristic of individual females.

Dr. B. P. Sonnenblick kindly supplied a different *Ore R* stock, which was selected several years ago at New York University for high egg-hatchability and which will be referred to as *Ore R (S)*. The failure of egg-hatching is much less frequent in *Ore R (S)* than in other stocks, and averages only about 4%. When eggs are collected from single females of this stock, the failures of egg-hatching among the offspring of different females are, however, still non-random. Successive egg collections from the same female seem to indicate that in this case also a lower rate of hatching is not accidental but related to biological individual factors. Failures of egg-hatching, however, only rarely exceed 15% among the offspring of any individual female.

The following reciprocal matings were then carried out: (a) *Ore R (S)* ♀ by *Ore R (S)* ♂, (b) *Ore R (S)* ♀ by *Sw b* ♂, (c) *Sw b* ♀ by *Sw b* ♂, (d) *Sw b* ♀ by *Ore R (S)* ♂. After one day, 15 females were isolated from each mating, and hatchability was tested on two successive samples of eggs from each female. (Out of 60 tests, 2 were discarded because no egg was laid, and 1, belonging to group (c), because all eggs lacked chorion and failed to hatch.) The observed numbers of hatched and unhatched eggs for the different groups were: (a) 857:49, (b) 692:26, (c) 868:129, (d) 732:179. Failures of hatching never exceeded 25% for any one female of groups (a) or (b), but exceeded 50% for one female of group (c) and for three females of group (d). This suggests that the non-random failures of hatching are largely determined by the female parent.

The eggs laid by one of the "semi-sterile" females of group (d) were allowed to develop and to undergo free brother-sister mating. Eggs were then collected from 11 single  $F_1$  females and tested. The total numbers of hatched and unhatched eggs were 843 and 6. The conditions which were responsible for the "semi-sterility" of the *Sw b* parental female thus ap-

peared to be wholly recessive, if inheritable at all. (However, later tests on the offspring of several  $F_1$  females, hybrids of *Ore R (S)* and *Sw b*, showed that the incidence of failures in egg-hatching is not lower, on the average, for these hybrid females than for inbred *Ore R (S)* females.)

Six strains of  $F_2$ , arising from as many  $F_1$  females out of the group of 11 referred to above, were allowed to develop and were mated by brother-sister pairs. Approximately 20 females from each strain were tested for egg-hatchability. Eggs were collected for one day from 117 females; 8 tests were discarded because no eggs or only 1 or 2 eggs were collected, while all other females laid at least 10 eggs. The results are shown in table 1. The frequency distribution in the last row is clearly bimodal, indicating segregation of some major factor. If 50% is taken as the abscissa separating the two modes, the total frequencies within each mode are 82 and 27; that is, they are approximately in the ratio 3:1, indicating the probable existence in the parental female of a single major mendelian factor. The frequency distribution is, however, not quite homogeneous among different strains; this should be related to minor factors which were heterozygous in the parental generation.

TABLE 1

TESTS OF THE FERTILITY OF THE  $F_2$  OF A SINGLE "SEMI-STERILE" *Sw b P* ♀

NUMBER OF FEMALES CLASSIFIED ACCORDING TO THEIR FERTILITY														
STRAIN NO.	PERCENTAGE OF UNHATCHED EGGS											TOTAL NUMBER OF ♀'S TESTED	TOTAL NUMBER OF EGGS OBSERVED	
		0	10	20	30	40	50	60	70	80	90		HATCHED	UN- HATCHED
	0	TO 10	TO 20	TO 30	TO 40	TO 50	TO 60	TO 70	TO 80	TO 90	TO 100			
1	1	5	1	..	1	..	1	1	2	2	..	14	290	150
2	5	10	1	1	..	1	..	1	..	..	..	19	723	89
5	10	7	..	1	..	..	..	..	1	..	1	20	579	81
6	6	6	1	2	1	1	..	1	..	1	1	20	563	182
7	6	5	..	..	..	..	2	3	2	1	..	19	428	181
8	3	7	..	..	..	..	1	4	1	..	1	17	450	150
Total	31	40	3	4	2	2	4	10	6	4	3	109	3033	833

The  $F_3$  from the 27 "semi-sterile"  $F_2$  females which had shown over 50% failure of egg-hatching was further tested for fertility. This was actually a test to determine which  $F_2$  males had been homozygous for the same major factor as the females to whom they were mated. Beginning at this stage of the experiments, the classification of a female as "semi-sterile" was arbitrarily determined according to whether the percentage of failures in egg-hatching did or did not exceed 50%. In the large majority of cases an actual egg-count was not carried out, and the classification could easily be performed by glancing at a batch of eggs after incubation. All females within 6 of the 27 strains of  $F_3$  appeared to be "semi-sterile." These 6 strains were carried further; inbreeding of each strain and cross-breeding between strains produced only "semi-sterile" females, indicating

that a single "semi-sterility" factor—thereafter tentatively called *sst*—had been isolated.

The segregation experiments reported above made it appear likely that *sst* is autosomal. The usual test for location of a gene was then carried out: *sst/sst* males were mated to *Pm/Cy H/Sb* virgin females, and 10  $F_1$  *Cy Sb* males were back-crossed singly to *sst/sst* virgin females. The phenotypically different classes of  $F_2$  females were tested for *sst*, with the following results: of 72 non-Curly females tested, 57 produced offspring and all were homozygous for *sst*; 95 out of 107 tests on *Cy* females were successful, and all but 2 of them were clearly not "semi-sterile." *Sb* appeared to be wholly unrelated to *sst*. Therefore, *sst* is located in the second chromosome. The nature of the exceptions represented by the two "semi-sterile" *Cy* females is still under investigation. These flies were the offspring of two different  $F_1$  males.

*Discussion.*—The constitutional characteristics of a female affecting the viability of its offspring might have been expected to be determined by the collective action of a large number of factors, each of them only slightly important in itself. The probable identification of the factor *sst* indicates that this is not so.

Several questions which may be of some interest, particularly to the physiological geneticists, can now be considered. First, the way in which *sst* acts can be investigated. The failures of development in the offspring of *sst* females probably occur at an early stage of the egg development because the unhatched eggs do not show any obvious sign of embryonic development when dechorionated. It is possible either that the eggs themselves are defective or that *sst* females do not succeed in utilizing properly the sperm they have received.

Further investigation might also be directed toward determining how widespread the factor *sst* is among several laboratory stocks. It is not yet clear how this factor survived throughout a long series of experiments prior to its identification, in spite of its great selective disadvantage.

Finally, the problem of controlling the failures of development is not completely solved by eliminating *sst*. Females other than *sst*, whose offspring show embryological disturbances to a high degree, have frequently been found. Random presence of such females in any experimental sample is liable to have a disturbing effect. The impression has been gathered that many more genetic factors inducing failures of development to a more or less striking degree could be easily identified. Investigation of such factors would be especially interesting if they appeared to act through the production of defective eggs, lacking, for instance, in some growth factor.

The "semi-sterility" of the two exceptional *Cy* females may have been due to factors other than *sst*, although the ratios of segregation (1:17 and 1:11) appear to be rather small. If, however, the factor *sst* was actually

present in the same chromosome as *Cy*, it ought to have arisen by mutation, since no crossing-over occurs in males.

*Summary.*—A wild type strain, most of whose eggs fail to develop, has been isolated. Breeding experiments indicate that this behavior is due to a second-chromosome factor which affects the offspring of homozygous parent females independently of the genetic constitution of their mates.

<sup>1</sup> Fano, U., and Demerec, M., *Genetics*, 26, 151 (1941).

---

## "ATAXIA," A HEREDITARY NERVOUS DISORDER OF THE RABBIT

BY P. B. SAWIN, M. V. ANDERS AND R. B. JOHNSON

DEPARTMENT OF BIOLOGY, BROWN UNIVERSITY

Communicated March 11, 1942

Many degenerative disorders of the nervous system in man occur more often in certain families than in the general population, suggesting that inheritance is involved, but conclusive evidence for this is difficult to obtain.

But in laboratory animals produced in large numbers under uniform conditions and inbred as closely as is practicable, evidence as to the inheritance of nervous disorders is readily obtained. We have observed such a case in the rabbit colony at Brown University, in a family of chinchilla rabbits which has been under continuous observation for nearly 15 years, and in which parent-offspring inbreeding had been practiced for at least 5 generations at the time the disorder was first discovered. The disorder is one which for lack of a more precise descriptive name we have called "ataxia." It is due to a semi-lethal mutation which is inherited as a simple recessive character. The genetic symbol to be used in reference to this mutation is *at*, its normal and dominant allele being *At*.

The first affected animal was observed late in the winter of 1936, in a litter of six, all of which appeared vigorous and entirely normal at birth. At two months of age one of the females became paralyzed in the hind quarters. She was kept under observation but died after a period of two weeks during which time the condition became progressively worse. Each of the other two females of this litter remained entirely normal, was raised to maturity and mated back to the father. Each produced young in her first litter which showed an identical condition setting in at approximately the same age. This gave rise to the suspicion that the defect might be due to a hereditary mutation.

Since affected individuals do not survive until sexually mature, all which occur are born of normal parents. Accordingly the mutation must be re-

cessive in inheritance. When it arose cannot be stated since both parents of the first affected individual must have carried the mutant gene derived from some common ancestor.

The first ten individuals which manifested the defect were progeny of the same male (4207), a vigorous normal rabbit still sexually active although seven years old. He produced "ataxic" young by two females of quite different immediate ancestry, one belonging to his own line of descent, the other derived from an outcross of his ancestor (♂48D3) with an entirely different line of chinchilla rabbits. Accordingly the mutation must have been already in existence as a recessive in the nearest common ancestor of both females, viz., ♂48D3.

The defect manifests itself first in the hind legs as a disinclination to movement and secondly in the manner of movement which is forced and irregular. Instead of the characteristic hopping movements of normal rabbits with both hind legs working together, there is a tendency for them to be used alternately and with seeming lack of coördination. In the next stage one of the hind legs appears to be stiffened and extended to one side. The weight of the body is then carried by the other leg, producing a posture which is quite characteristic. Even before this can be readily recognized a tendency may be noted for the animal to come to rest close to the side of the cage, as if for support.

The gait becomes progressively clumsier, the legs gradually lose their ability to maintain support of the body and the animal, after walking a short distance, may fall over on its side in an apparently helpless condition. After a rest, it may be able again to utilize its legs in support of the body. In from 5 to 10 days the condition may progress to the point where the animal is almost completely incapacitated. At this stage the hind legs can never be used to support the body. The animal lies upon its side, usually the one opposite the leg which first became affected, with the hind legs extended. When the animal is disturbed, walking movements may be attempted but exhaustion comes rather quickly.

At this stage the animal seems hypersensitive to many situations which are not disturbing to a normal rabbit. Although sudden movements and loud noises do not cause violent reactions, the animal does appear momentarily startled. Continued stimulation such as caused by people moving about the room leads to restlessness which may continue long after the stimulus has been removed. This hypersensitivity is also strikingly manifested in the efforts of the animal to drink from the customary brass nipple and water bottle employed in our laboratory. The vibrissae appear so highly sensitive that any approach to the nipple causes violent withdrawal and may account in part at least for the very apparent resultant dehydration of the animals at autopsy. It is also interesting to note that the animal appears less sensitive to glass than to brass nipples. At about

this time a characteristic grinding of the teeth, which continues at irregular intervals until death, becomes apparent.

The appetite appears to differ with the individual. Some animals lose appetite very soon after onset of the condition. Others have shown a desire for food for some time after they were no longer able to gain their feet, and therefore had to be fed by hand. Food is located with difficulty even if placed in close proximity to the head. In general it has been found that succulent green food will be taken long after all other food is refused and a better general condition is maintained as a consequence.

Sensitivity and excitability seem to increase to a definite peak which varies with the individual. The patellar reflex first shows movements of extremely low magnitude followed by increasing rapidity and magnitude until a crossed reflex is present. At no time does the patellar reflex appear pendular. This reflex degeneration appears to progress anteriorly until the fore legs and head are involved. A crossed reflex has never been observed in the fore leg although responses of low magnitude result followed by responses of high magnitude. At this stage some individuals show a pronounced lateral nystagmus. Breathing tends to be convulsive and orientation of head movements become difficult. Pain reflexes continue to be present until death, although some minor changes in them have been observed.

In the final stage, the animal becomes more or less flaccid, and in contrast to the preceding hypersensitivity, becomes relatively quiescent with a marked drop in body temperature. It is not easily stimulated and apparently is unaware of its surroundings. All desire for food or water is lost. Respiration is more irregular and labored and death follows within a few hours.

Post-mortem examination usually shows excessive dehydration and emaciation even in those individuals having the best appetites. The tongue shows no abnormality such as is described for vitamin B deficiencies, with which certain types of paralysis are sometimes associated, nor is the coat in any way affected. The teeth, particularly the incisors, are frequently worn and chipped beyond those of normal sibs, which is not surprising in view of the continuous grinding already mentioned. The adrenals in those animals which have been examined seem to be smaller than is normal for rabbits of the same age although we are not able to attach any particular significance to this observation. Other post-mortem irregularities have not been observed.

The time of first manifestation of the condition has varied from 58 to 89 days after birth, the mean for 41 individuals being 72.9 days. The true mean may be earlier since the earliest indications of onset could easily be overlooked. The sequence of the above described events has been the same in all cases observed, each stage differing only in duration and inten-



sity. The longest duration of the condition was observed in a case lasting 27 days, death probably being due to involvement of the respiratory centers. Several have succumbed five days after first manifesting symptoms. Preliminary microscopic examination of the nervous system has revealed demyelinated areas in the medulla in the general vicinity of the vestibular nuclei. Although we cannot as yet state the exact extent or variability of the lesions, it seems probable that vestibulo-cerebellar and ponto-cerebellar systems are involved. These observations as well as the symptoms and reflex behavior described above indicate that we are dealing with an extrapyramidal degenerative disorder.

Inasmuch as the affected rabbits are always derived from entirely normal appearing parents, and die before reaching sexual maturity, the character may be described as a sublethal recessive. It can be perpetuated and studied only by way of the heterozygous carriers. Matings between such known heterozygous males and females have produced a total of 219 young which have been reared to at least three months of age. The affected individuals number 55, the expectation for a monofactorial recessive being 54.75; hence there is no prenatal lethality of the mutation.

The use of extracted recessives in the customary outcross, back-cross and *inter se* recessive matings being impossible, known heterozygotes were used in their stead. In 12 out-cross matings involving two different unrelated breeds, family X descended from Castle's small race and family III a New Zealand White, 64 normal young have been obtained. The subsequent back-cross matings of these  $F_1$ s to known heterozygotes of family V appear to fall in two categories, those which produce affected young and those which do not. Six or more normal young have been regarded as an adequate test. Of such tested  $F_1$ s, 6 were heterozygous ( $At\ at$ ) and 4 were homozygous ( $At\ At$ ), equality being expected.

Of the normal individuals derived from heterozygous parents,  $1/3$  should be homozygous and therefore incapable of producing affected young, whereas  $2/3$  should be heterozygous and so capable of producing affected young. Of 34 individuals tested, 12 proved to be homozygous and 22 heterozygous, which is close to the expected 1:2 ratio. Affected individuals occur in both sexes without statistically significant difference in frequency.

As in the case among pigeons described by Riddle<sup>1</sup> it is not certain by what name this disorder should be called. In the sense that it represents a loss of coördination of movements it is an ataxia. However, neither in the symptoms manifested nor in the pathology described does the case of the pigeon seem to be strictly comparable with this. This disorder is more closely related to the condition described in the rabbit by Nachtsheim<sup>2</sup> as spastic spinal paralysis. But the time of onset of the present defect is much later than that reported by Nachtsheim and the duration is shorter. The pathological disturbance appears to be located in the same general re-

gions of the brain although it is difficult to be certain as to their location from the information available for the spastic paralysis. The inheritance in all of these cases is similar. The case of paralysis in the dog described by Stockard<sup>3</sup> is entirely dissimilar in symptomatology, pathology and also in inheritance.

The present disorder, due to its semi-lethal nature, is not ideal genetic material with which to work but genetic tests of the possible chromosomal association with other known genes have been devised which at the same time are providing an economical and continuous supply of affected animals with which to test the functional and organic nature of this condition, which so closely parallels certain human degenerations.

*Conclusion.*—We may conclude that this degenerative disorder involving the brain stem, which appears to have arisen by a mutation, is transmitted as a simple mendelizing unit. By crossing it can be introduced into unrelated and unaffected families to become manifest in subsequent generations. Likewise it can be eliminated by proper breeding tests accompanied by selection of those segregates which are free from the defect, for establishment of a new line.<sup>4</sup>

<sup>1</sup> Riddle, O., *Proc. Soc. Exper. Med. and Biol.*, 15, 56 (1918).

<sup>2</sup> Nachtsheim, H., *Deut. Tierärztl. Wochenschrift*, 44, 742-746 (1936).

<sup>3</sup> Stockard, C. R., *Am. Jour. Anat.*, 59, 1-54 (1936).

<sup>4</sup> This study is a portion of a research program supported in part by a grant from the Rockefeller Foundation.

---

## THE IDENTIFICATION AND CHARACTERIZATION OF BACTERIOPHAGES WITH THE ELECTRON MICROSCOPE

BY S. E. LURIA\* AND THOMAS F. ANDERSON†

BACTERIOLOGICAL RESEARCH LABORATORIES OF THE DEPARTMENT OF SURGERY,  
COLLEGE OF PHYSICIANS AND SURGEONS, COLUMBIA UNIVERSITY, NEW YORK, AND  
RCA RESEARCH LABORATORIES, CAMDEN, N. J.

Communicated March 12, 1942

Bacteriophages, or bacterial viruses, are a group of viruses reproducing in the presence of living bacterial cells. Bacteriophages are particulate, and convincing evidence exists that (1) one particle of phage is sufficient to originate the lysis of a bacterial cell; in the lysis, a variable number of new phage particles (an average of 100 or more) are liberated per cell;<sup>1</sup> (2) the elementary particles of each phage strain seem to have a characteristic particle size as determined by any one of various indirect methods of investigation (ultrafiltration,<sup>2</sup> radiosensitivity,<sup>3</sup> diffusion,<sup>4</sup>) and diameters ranging from 10 to 100 m $\mu$  have been obtained for the various strains de-



pending on the method of investigation, although diffusion experiments occasionally yield still smaller values.

The electron microscope has recently been applied with success to the study of viruses<sup>5</sup> and it therefore seemed desirable to attempt such a study of bacterial viruses, particularly since they offer favorable possibilities for the identification of the virus particles through a study of the reaction between the individual particles and the bacterial cell under the microscope. Indeed, a number of short reports have been published recently by German authors<sup>6, 7</sup> in which round particles have been described as corresponding to the phage particles, although Ruska<sup>7</sup> shows pictures of "sperm-shaped" particles from a phage suspension adhering to a bacterial membrane. From this evidence alone he is unable to decide whether these are particles of phage or bacterial products.

We have undertaken an investigation of the problems of phage structure, size, reproduction and lytic activity by means of the RCA electron microscope. Research on the last items is still in progress. The present report concerns itself with the identification and the morphological analysis of a number of strains of phage particles and their adsorption on sensitive bacterial cells. The results are illustrated by some of the electron micrographs (Plates I and II) which have brought to light many extremely interesting features. Details of the material and methods used will soon be published.

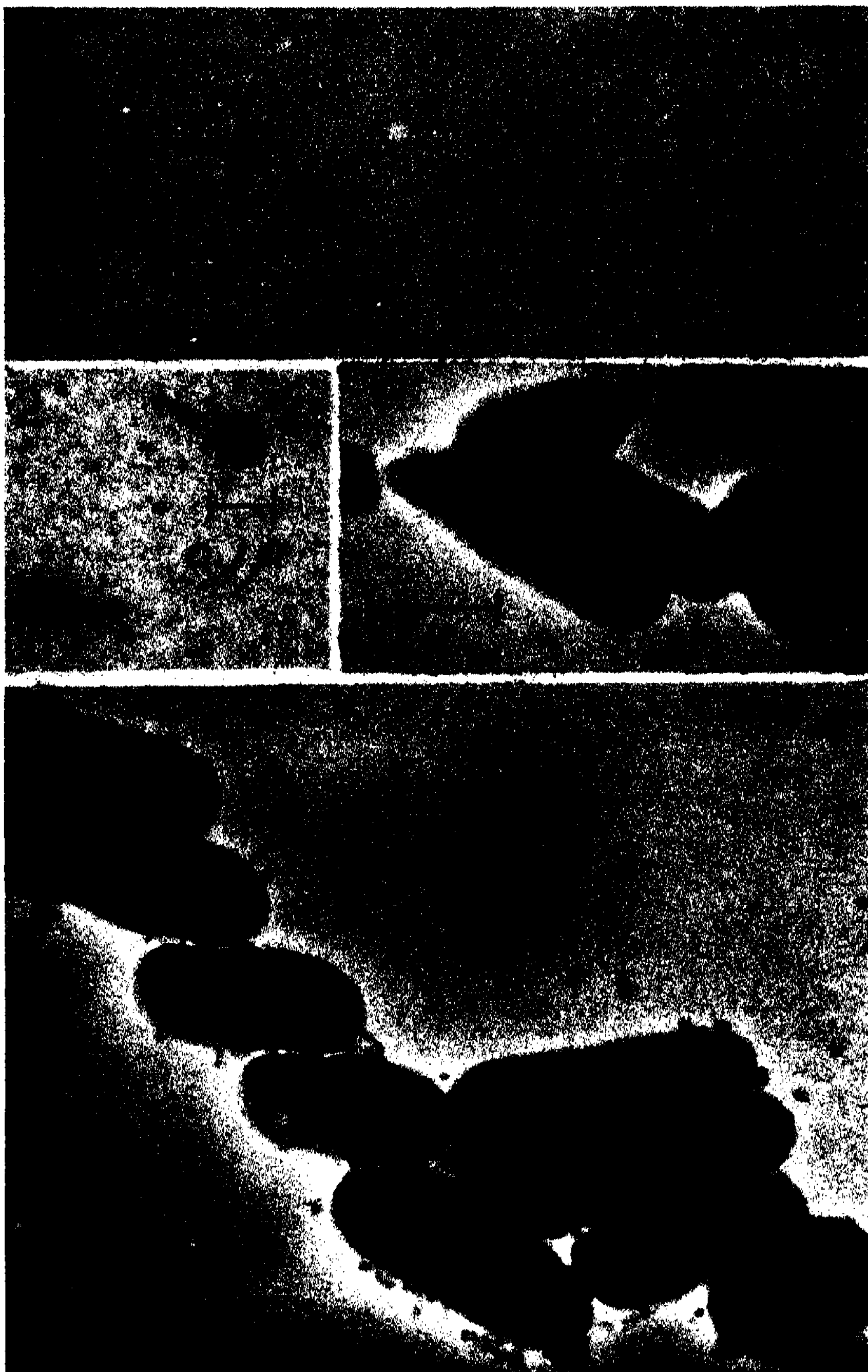
*I. Bacteriophage anti-coli PC* (particle diameter by diffusion 44 m $\mu$ , Kalmanson and Bronfenbrenner<sup>8</sup>; by x-irradiation 50 m $\mu$ , Luria and Exner, unpublished).

Micrographs of high titer suspensions, figures 1, 2, 4, 5 and 6, show the constant presence of particles of extremely constant and characteristic aspect. They consist of a round "head," and a much thinner "tail," which gives them a peculiar sperm-like appearance. The "head" is not homogeneous but shows an internal structure consisting of a pattern of granules, distinguished by their higher electron scattering power. Deviations from the usual symmetrical internal pattern may be due to varying orientation of the particles or to other factors as yet unknown. The diameter of the head appears to be about 80 m $\mu$ ; the tail is about 130 m $\mu$  long.

#### EXPLANATION OF PLATE

##### PLATE I

1. Electron micrograph of particles from a high titer suspension of bacteriophage anti-coli PC.  $\times 38,000$ .
2. Particles from a high titer suspension of bacteriophage anti-coli PC.  $\times 84,000$ .
3. *Escherichia coli* from suspension in distilled water.  $\times 17,000$ .
4. *Escherichia coli* in suspension of bacteriophage anti-coli PC for ten minutes.  $\times 17,500$ .





This gives a size which is in fair agreement with the figures deduced from the radiosensitivity method. On the other hand, it is possible that the size as determined by x-rays corresponds more closely to the size of the granules.

When allowed to stand a few minutes in the presence of sensitive bacterial cells *Escherichia coli*, strain PC (Fig. 3), the particles described above are readily adsorbed (Figs. 4 and 5). They appear to stick to the bacteria either by the head or by the tail. Other conditions remaining constant, the number of particles adsorbed on a bacterium increases with the time of contact, although it is difficult at the present time to differentiate between adsorption and reproduction of the particles on the cell wall. By allowing the phage to stay in contact with bacteria for a time of the order of the minimum time of lysis (21 minutes for PC phage, Delbrück and Luria<sup>1</sup>) it is possible to observe bacterial cells extensively damaged, surrounded by a very large number of particles, probably newly formed (Fig. 6).

II. *Bacteriophage anti-coli P 28*, also active on *Escherichia coli* strain PC (particle size: irradiation, 36 m $\mu$ , Luria and Exner.<sup>3</sup>

Round particles are visible in the suspensions of this phage which are somewhat smaller than those described for phage PC (about 50 m $\mu$  in diameter). An extremely thin tail, although difficult to demonstrate with certainty in the reproductions, seems to be visible in many instances. In many micrographs the head is almost completely filled by a dense internal structure. These particles, too, are readily adsorbed on sensitive bacterial cells.

III. *Bacteriophage anti-staphylococcus 3K* (particle size: by ultrafiltration\* and ultracentrifugation 50–75 m $\mu$ , Elford;<sup>2</sup> by irradiation 48 m $\mu$ , Luria and Exner.<sup>3</sup>

Owing to technical reasons, the conditions for successful micrographing are here less favorable. Nevertheless, the presence of approximately round particles of proper size has been established in preparations of this phage also.

We are inclined to identify the particles described above with the actual particles of bacteriophage for the following reasons: (a) They are always present in highly active phage suspensions and missing in any control suspensions (media, bacterial cultures, bacterial filtrates, etc.); (b) they are readily adsorbed by the bacterial cells of the susceptible strain and fail to be adsorbed by other bacteria; (c) the size from a given strain is uniform and corresponds essentially to measurements by indirect methods; (d) the structure of both the "head" and the "tail" is characteristic of the strain of phage; (e) preliminary experiments on the lysis process seem to demonstrate the liberation of these particles from the lysing bacteria.

*Conclusions.*—We do not want to discuss here the bearing of the above described results on the problem of the nature of bacteriophage and of viruses in general. We limit ourselves to pointing out the extreme interest of the finding of such constant and relatively elaborate structural differen-

tiation in objects of supposedly macromolecular nature. This result is of equal interest in the field of genetics, since genes, together with viruses, are currently supposed to be macromolecular entities.

The correspondence of the particle size as directly portrayed in the electron microscope with the results of indirect methods is also very remarkable, although it does not exclude the possibility of phage activity being sometimes associated with smaller particles. It is worth while emphasizing that the results of the present investigation, together with the recently published results of irradiation of bacteriophages, represent most desirable evidence for the validity of the so-called "hit theory" for the determination of the "sensitive volume" in sub-light-microscopic biological objects. This conclusion, too, seems to be interesting from the point of view of genetics, since the "hit theory," although widely criticized, has been used for calculating the approximate value of the dimensions of genes.

The authors are grateful to the National Research Council Committee on Biological Applications of the Electron Microscope for allocating time for this study, and to the RCA Laboratories for the use of their facilities, and to Dr. V. K. Zworykin for his interest and encouragement. The authors also thank Dr. Stuart Mudd for the use of the facilities of his laboratory for the preparation of material for study.

\* Aided by a grant from the Dazian Foundation for Medical Research.

† RCA Fellow of the National Research Council.

<sup>1</sup> Ellis, E. L., and Delbrück, M., *Jour. Gen. Physiol.*, **22**, 365 (1939); Delbrück, M., and Luria, S. E., to be published.

<sup>2</sup> Elford, W. J., in Doerr and Hallauer, *Handbuch der Virusforschung*, vol. I, Julius Springer, Wien, 1938, p. 126.

<sup>3</sup> Luria, S. E., and Exner, F. M., *Proc. Nat. Acad. Sci.*, **27**, 370 (1941).

<sup>4</sup> Hetler, D. M., and Bronfenbrenner, J., *Jour. Gen. Physiol.*, **14**, 547 (1931).

<sup>5</sup> Stanley, W. M., and Anderson, T. F., *Jour. Biol. Chem.*, **139**, 325-538 (1941), and references given therein.

<sup>6</sup> Pfankuch, E., and Kausche, G. A., *Naturwiss.*, **28**, 46 (1940).

<sup>7</sup> Ruska, H., *Naturwiss.*, **29**, 367 (1941).

<sup>8</sup> Kalmanson, G. M., and Bronfenbrenner, Jr., *Jour. Gen. Physiol.*, **23**, 203 (1939).

---

## EXPLANATION OF PLATE

### PLATE II

5. *Escherichia coli* in suspension of bacteriophage anti-coli PC for 20 minutes. X 14,500.

6. *Escherichia coli* in suspension of bacteriophage anti-coli PC for 20 minutes. X 12,500.

7 and 8. Particles from a high titer suspension of bacteriophage anti-coli P28. X 38,000.

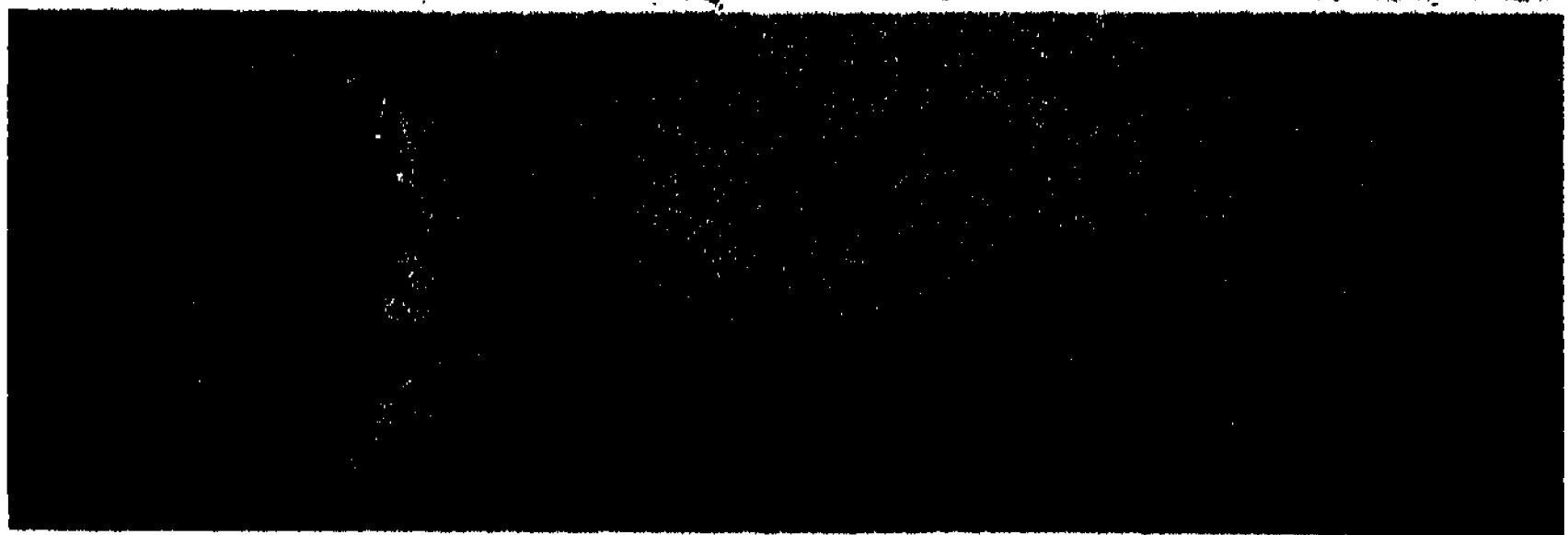


PLATE II



*CONCENTRATION OF GROWTH HORMONE AND FRUITFULNESS  
IN THE MONTMORENCY CHERRY*

BY FELIX G. GUSTAFSON

DEPARTMENT OF BOTANY, UNIVERSITY OF MICHIGAN

Communicated March 11, 1942

Several years ago the writer<sup>2</sup> discovered that varieties of oranges, lemons and grapes that produce seedless fruits have a higher concentration of growth hormone in the ovary of the flower than varieties that produce seeded fruits. This led to the interesting speculation as to whether there might be any relation between growth hormone content of the ovaries and fruitfulness in a species of plant. But no immediate material was available.

A visit to Prof. V. R. Gardner, Director of the Michigan State Agricultural Experiment Station, brought forth the information that the Station had selections of fruitful and barren Montmorency cherries. These trees were put at the disposal of the writer, and it was decided to set up an experiment to investigate the growth hormone content of these selections. This investigation was continued for three seasons. Three trees of a high yielding selection and three trees of the low yielding selection were chosen. The high yielding trees were nineteen years old and of a standard type; the bearing records of these trees were available for the whole period. The low yielding selections originated as a bud sport about 1923.<sup>1</sup> The trees chosen were nine years old; their yield record, and that of the tree from which they had been derived, was also known. For several years the percentage of setting had been determined for these various trees, and it was found that on the barren trees from 0.4 to 11.0 per cent of the blossoms produced fruits, whereas the fruitful trees set from 20.0 to 47.0%. These figures show that there is a great difference in the percentage of setting in the two lots of trees. Gardner<sup>1</sup> has examined numerous ovaries from the barren trees, and has found that they are perfectly normal. He states that the flowers are somewhat larger than those on the fruitful trees and the writer has noticed that too, as well as that the ovaries are larger. Every year that the trees have been examined, it was observed that the barren trees produced a profusion of blossoms.

The procedure followed in the experiment was to collect blossom buds that were ready to open or were in the process of opening. By two cuts, the ovary was separated from the bud. The ovaries were placed in closed vials until a number had been obtained, and during the first two seasons these ovaries were placed in freshly distilled ether as soon as weighed, but in 1941 they were frozen with dry ice as soon as dissected out, and were kept frozen until reaching Ann Arbor, where they were boiled for one minute



before placing in ether.<sup>3</sup> The further extraction and auxin assay were made in the usual way.

Material for the first analysis was collected on May 9 and 10, 1939. Three different collections of flowers were made during these two days, and each time flowers were gathered from both groups of trees so that there would be no difference in the weather at the time of collection. As shown by table 1, the growth hormone concentration was approximately 23% greater in the ovaries from the fruitful trees than from the barren trees. In 1940, the flowers were collected from the barren trees in the afternoon and evening of a very warm and sunny day, and collections were made from the fruitful trees next morning and afternoon during a cold rain. This time the barren trees showed a slightly higher growth hormone content.

TABLE 1

RELATION BETWEEN AUXIN CONCENTRATION IN THE OVARIES AND FRUITFULNESS IN THE MONTMORENCY CHERRY. AUXIN CONCENTRATION IS EXPRESSED AS INDOLE ACETIC ACID EQUIVALENTS IN GAMMAS PER KILOGRAM OF FRESH MATERIAL.

YEAR	FRUITFUL TREES			UNFRUITFUL TREES		
	AMOUNT OF MATERIAL, IN GRAMS	AUXIN CONCENTRATION GAMMAS PER KILOGRAM	YIELD OF FRUITS, IN POUNDS	AMOUNT OF MATERIAL, IN GRAMS	AUXIN CONCENTRATION GAMMAS PER KILOGRAM	YIELD OF FRUITS, IN POUNDS
1939	3.445	20.07	770	2.390	16.27	2.12
1940	9.760	16.57	321	4.125	17.92	Only a few fruits
1941	4.456	24.90	430	3.390	8.02	61.0

NOTE: Three trees were used of each kind.

The results for the two years left the matter in a very uncertain situation. While some unpublished data seem to indicate that growth hormone content is higher in a plant in a warm temperature than in a cold one, the information is not too positive. If that were the situation, we might be able to explain the low concentration in the fruitful trees and the high one in the barren trees. To avoid any such complications, collections were made from all trees at the same time in 1941. Both days that the material was gathered this year the weather was very warm and the days were sunny. This year, the fruitful trees had a growth hormone content more than three times as great as that of the barren trees. The experiment in 1941 was the most carefully executed of all. The new method<sup>3</sup> of freezing and destroying the enzyme system by boiling has been found to give results more indicative of the true values of active growth hormone in a plant than the older method. This year greater care was also exercised in the selection of the flower buds, and in the time of collection.

The results seem to the writer to justify the conclusion that in the Montmorency cherry there is a positive correlation between fruit setting and the growth hormone content of the ovary of the flower. But before

one can draw any general conclusions, more comparisons between barren and fruitful plants would have to be made.

The writer wishes to express his gratitude to Professor V. R. Gardner for his aid in this investigation, and also to the Directors of the Horace H. Rackham Research Fund for financial aid.

<sup>1</sup> Gardner, V. R., "A Study of Some Unproductive Sports of the Montmorency Cherry," *Jour. Agri. Res.*, 50, 457-478 (1935).

<sup>2</sup> Gustafson, F. G., "The Cause of Natural Parthenocarpy," *Amer. Jour. Bot.*, 26, 135-138 (1939).

<sup>3</sup> Gustafson, F. G., "The Extraction of Growth Hormones from Plants," *Ibid.*, 28, 947-951 (1941).

---

## THE CALCULATION OF ROTATION FACTORS FOR ECLIPSING BINARY SYSTEMS

BY ZDENĚK KOPAL

HARVARD COLLEGE OBSERVATORY

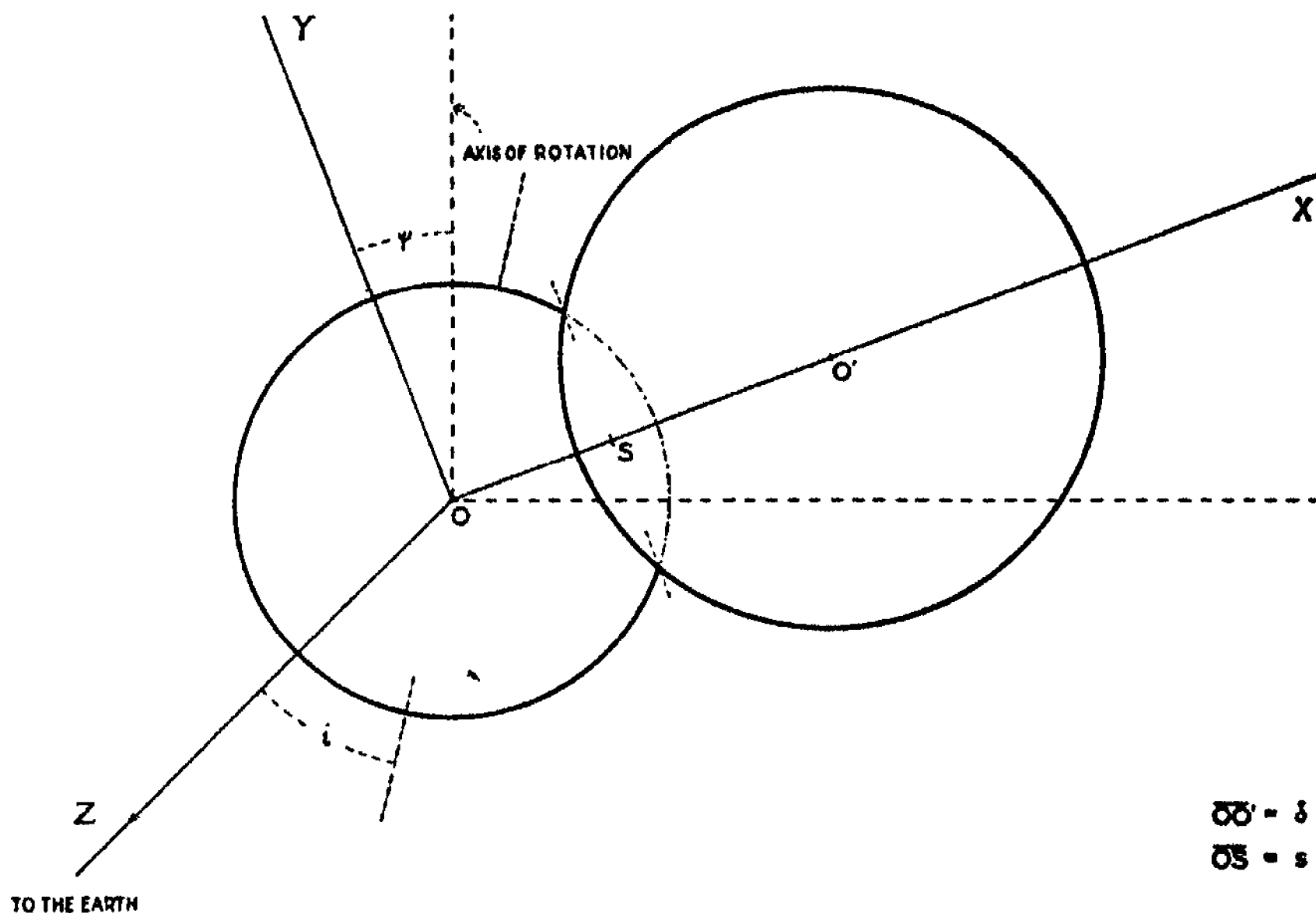
Communicated March 16, 1942

The interpretation of the rotational effect, exhibited by the velocity curves of eclipsing systems within minima, appears to be one of the few possible methods to specify the absolute dimensions of single-spectrum binaries. A first step of such a procedure is to determine the velocity of rotation  $V_e$  at the equator of the star undergoing eclipse. If  $R$  denotes the observed rotational effect, i.e., the deviation of observed radial velocity of the eclipsed star from a simple elliptic motion in a plane inclined to the celestial sphere by an angle  $i$ , then the ratio

$$R/V_e \sin i = F \quad (1)$$

defines a dimensionless quantity which depends only on the geometry of the eclipse and on the distribution of brightness over the apparent disc of the eclipsed star. The quantity  $F$  defined by the foregoing relation is known as the rotation factor. Its value varies with the phase, but can evidently be predicted for any particular moment provided that the geometrical elements of the eclipse are known from the study of a light curve. The aim of the present paper will be to establish analytical formulae expressing  $F$  explicitly in terms of such elements for stars which appear uniformly bright or are darkened at limb to any arbitrary degree. Formulae relevant to uniform discs have already been set up by Petrie;<sup>1</sup> but his results will be found to admit of considerable simplification; while the darkened case has so far not been tackled analytically at all.

First, let us define  $F$  in terms of the geometry of eclipse. Provided that the star rotates like a rigid body, the radial velocity at any point on the apparent disc—apart from any orbital motion—is proportional to the distance  $l$  of that point from the projected axis of rotation. Let us refer our eclipsing system to a rectangular frame  $xyz$  (see the figure), with origin at the center of component undergoing eclipse, defined so that the  $xy$ -plane is tangent to the celestial sphere and the  $z$ -axis coincides with the line



of sight, while the  $x$ -axis is constantly in the direction of the projected center of the secondary component. Then, evidently,

$$F = \frac{\int \int J l dx dy}{r_1 \int \int J dx dy}, \quad (2)$$

where  $r_1$  is the radius of the star undergoing eclipse and  $J$  specifies the distribution of brightness over its apparent disc. The limits of integration are to be extended over the whole area visible at any particular phase.

Let, further,  $\theta$  denote the longitude in the plane of a circular orbit, reckoned from the mid-primary minimum, and  $\delta$  be the apparent separation of centers of both components expressed in terms of the orbital radius taken as unit. The distance  $l$  of a point, on the apparent disc, from the projected axis of rotation then becomes

$$l = x \cos \psi - y \sin \psi, \quad (3)$$

where  $\psi$ , the angle between the axis of rotation projected on the celestial sphere and the  $y$ -axis, follows from

$$\delta \cos \psi = \sin \theta \quad \text{and} \quad \delta \sin \psi = \cos \theta \cos i,$$

while

$$\delta^2 = \sin^2 \theta \sin^2 i + \cos^2 i.$$

As to  $J$ , the distribution of brightness over the apparent discs of spherical stars is governed by limb-darkening only, and hence

$$J = H(l - u + uz), \quad (4)$$

where  $H$  denotes the intensity of radiation parallel to the line of sight;  $u$ , the coefficient of limb-darkening; and  $z$  is the cosine of the angle of foreshortening which, in terms of the  $xyz$ -coördinates, is

$$r_1 z = \sqrt{r_1^2 - x^2 - y^2}.$$

The above formula for the rotation factor then takes the form

$$F = \frac{\cos \psi \iint J x dx dy - \sin \psi \iint J y dx dy}{r_1 \iint J dx dy}. \quad (5)$$

Since, however,

$$\iint J y dx dy$$

integrated over any area symmetrical with respect to the  $x$ -axis vanishes, equation (5) immediately reduces to

$$F = \frac{\sin \theta}{\delta r_1} \left\{ \frac{r_1(1 - u) \iint x dx dy + u \iint x \sqrt{r_1^2 - x^2 - y^2} dx dy}{r_1(1 - u) \iint dx dy + u \iint \sqrt{r_1^2 - x^2 - y^2} dx dy} \right\}. \quad (6)$$

In setting up the limits of integration, which has to be extended over the whole area visible at any particular phase, we find it convenient to perform the integration first over the whole disc, then over the eclipsed area, and subtract. The advantage of so doing rests with the fact that the expressions in the numerator on the right-hand side of (6) vanish if integrated over the whole disc, while integrals in the denominator reduce to  $\pi r_1^2$  and  $2/3 \pi r_1^3$ , respectively. We are then left with integrals extending over the eclipsed area, of the form

$$\left\{ \int_s^{r_1 + \sqrt{r_1^2 - x^2}} \int_{-\sqrt{r_1^2 - x^2}}^{+\sqrt{r_1^2 - x^2}} + \int_{\delta - r_2 - \sqrt{r_2^2 - (\delta - x)^2}}^s \int_{-\sqrt{r_2^2 - (\delta - x)^2}}^{+\sqrt{r_2^2 - (\delta - x)^2}} \right\} x^m z^n dx dy = \frac{x^m z^n dx dy}{\pi r_1^{2+m+n} \alpha_n^m}, \quad (7)$$

where  $r_2$  denotes radius of the secondary (eclipsing) component, and

$$s = \frac{r_1^2 - r_2^2 + \delta^2}{2\delta}.$$

The dimensionless quantities  $\alpha_n^m$  defined by equation (7) are evidently identical with the associated  $\alpha$ -functions recently investigated by the writer<sup>2</sup> in connection with effects of gravity-darkening upon light curves of close eclipsing binaries. The rotation factors rewritten in terms of the associated  $\alpha$ -functions take the simple forms

$$F = -\frac{\sin \theta}{\delta} \left\{ \frac{(1-u)\alpha_0^1 + u\alpha_1^1}{(1-u)(1-\alpha_0^0) + u(\frac{2}{3} - \alpha_1^0)} \right\}, \quad (8)$$

the negative sign being due to the fact that the formulae were set up for a phase after the mid-primary minimum when the rotational effect is negative.<sup>3</sup> For a corresponding phase preceding maximum eclipse the factors are numerically the same, but of opposite sign.

The evaluation of the associated  $\alpha$ -functions for various values of  $m$  and  $n$  has recently been completed by the writer,<sup>2</sup> so that only the final forms of functions involved in equation (8) need be quoted here. If  $n$  is zero (or an even integer) the respective associated  $\alpha$ -functions can be expressed in terms of circular and algebraic functions. In particular, for  $n = 0$  we have

$$\pi r_1^2 \alpha_0^0 = r_1^2 \cos^{-1} \frac{s}{r_1} + r_2^2 \cos^{-1} \frac{\delta - s}{r_2} - \delta \sqrt{r_1^2 - s^2}, \quad (9.0)$$

and

$$\pi r_1^3 \alpha_0^1 = \delta r_2^2 \cos^{-1} \frac{\delta - s}{r_2} - \delta(\delta - s) \sqrt{r_1^2 - s^2}, \quad (10.0)$$

if the eclipse is partial (regardless of whether the component undergoing eclipse is the larger or the smaller of the two), and

$$r_1^2 \alpha_0^0 = r_2^2, \quad (9.1)$$

$$r_1^3 \alpha_0^1 = \delta r_2^3, \quad (10.1)$$

if it is annular.

The functions  $\alpha_0^0$  and  $\alpha_0^1$  specify the rotation factors for uniformly bright stars alone.<sup>4</sup> If, however, there is some limb-darkening, two further associated  $\alpha$ -functions with  $n = 1$  ( $m = 0$  and  $1$ ) are required for complete specification of our problem; and these are expressible only in terms of

elliptic integrals. Integrating (7) for the respective values of  $m$  and  $n$ , after some tedious algebra we obtain

$$\pi r_1^3 \alpha_1^0 = \frac{2}{3} \mathfrak{E} - \sqrt{\frac{\delta}{2}} \{ \mathfrak{P}_1^0 \omega_1 - \mathfrak{Q}_1^0 \eta_1 \}, \quad (11)$$

and

$$\pi r_1^3 \alpha_1^1 = - \sqrt{\frac{\delta}{2}} \{ \mathfrak{P}_1^1 \omega_1 - \mathfrak{Q}_1^1 \eta_1 \}. \quad (12)$$

In these equations

$$\begin{aligned} \frac{1}{2} \mathfrak{E} &= \frac{\pi}{2} - \left\{ E\left(\frac{\pi}{2}, \kappa\right) - F\left(\frac{\pi}{2}, \kappa\right) \right\} F(\phi, \kappa') - \left\{ E(\phi, \kappa') - \frac{1}{2} \sqrt{\frac{\delta}{r_2}} \right\} F\left(\frac{\pi}{2}, \kappa\right), \\ \omega_1 &= \frac{1}{\sqrt{e_1 - e_3}} F\left(\frac{\pi}{2}, \kappa\right), \quad \eta_1 = \sqrt{e_1 - e_3} E\left(\frac{\pi}{2}, \kappa\right) - e_1 \omega_1, \end{aligned}$$

where  $F(1/2\pi, \kappa)$  and  $E(1/2\pi, \kappa)$  denote the Legendre complete integrals of the first and second kind with the modulus

$$\kappa^2 = \frac{e_2 - e_3}{e_1 - e_3},$$

while the respective incomplete integrals possess an amplitude of

$$\phi = \sin^{-1} \sqrt{\frac{2\delta}{r_1 + r_2 + \delta}}$$

and a complementary modulus

$$\kappa'^2 = \frac{e_1 - e_2}{e_1 - e_3}.$$

The constants  $e_1, e_2, e_3$ , involved in the foregoing relations, are defined by

$$e_1 = \frac{1}{3} (\delta - s) + r_2,$$

$$e_2 = -\frac{2}{3} (\delta - s),$$

$$e_3 = \frac{1}{3} (\delta - s) - r_2,$$

if the eclipse is partial, and

$$e_1 = -\frac{2}{3}(\delta - s),$$

$$e_2 = \frac{1}{3}(\delta - s) + r_2,$$

$$e_3 = \frac{1}{3}(\delta - s) - r_2,$$

if it is annular.

The coefficients of  $\omega_1$  and  $\eta_1$  in equations (11) and (12) take finally the forms

$$\mathfrak{P}_1^0 = \frac{2}{3} \left\{ g_3 + \frac{1}{12} (2\delta - s)g_2 + h^2(h - 2s) - r_1^2(h - 4s) \right\},$$

$$\mathfrak{P}_1^1 = \frac{1}{2} \left\{ \frac{1}{16}g_2^2 + \frac{1}{5}(8h - s)g_3 + \frac{1}{2}h(h - s)g_2 + (h^2 - r_1^2)(h^2 - 2hs + r_1^2) \right\},$$

$$\mathfrak{Q}_1^0 = \frac{4}{9} \left\{ 7r_2^2 - 4r_1^2 + \delta^2 \right\},$$

$$\mathfrak{Q}_1^1 = \frac{1}{2} \left\{ g_3 + \frac{1}{10}(14h - 3s)g_2 + 2(2h^2 - 3h^2s + r_1^2s) \right\},$$

where  $3h = 2\delta + s$ , and

$$g_2 = -4(e_1e_2 + e_1e_3 + e_2e_3),$$

$$g_3 = +4e_1e_2e_3.$$

If the ratio of radii  $k$  and the geometrical depth of the eclipse  $p$  are defined, as usual, by

$$k = \frac{r_s}{r_l} \quad \text{and} \quad p = \delta - \frac{r_l}{r_s},$$

where  $r_l$  and  $r_s$  denote the radii of the larger and smaller component, respectively, the associated  $\alpha$ -functions can evidently be made to depend on  $k$  and  $p$  alone and tabulated in terms of these parameters for any particular phase of the eclipse. In fact, for some of them, tables are already available. As the reader may easily verify,

$$\alpha_0^0 \equiv \alpha^U \quad \text{and} \quad \alpha_1^0 \equiv \frac{2}{3} \alpha^D,$$

where  $\alpha^U$  is the fractional loss of light due to the eclipse of a uniformly bright disc (equal to the fractional area eclipsed), and  $\alpha^D$  is a corresponding loss of light of a disc completely darkened at limb. Extensive and accurate tables of  $\alpha^U$  and  $\alpha^D$  have recently been completed by Zessewitsch.<sup>6</sup>

The denominators of (8) can therefore be found with the aid of tabular values with minimum of effort. No tables are, unfortunately, available so far for  $\alpha_0^1$  or  $\alpha_1^1$ . It is to be hoped that such tables will be constructed in a due time; but at present their values will have to be computed for every case using the formulae given above. The calculation of  $\alpha_0^1$  is simple; that of  $\alpha_1^1$  is likely to be rather tedious—yet much less so than a graphical computation of the rotation factors for darkened stars would be.

The formulae for the rotation factors set up in the present paper are exact as far as the components can be regarded as spherical—i.e., as long as their dimensions expressed in terms of their separation are small. In close binaries, however, the forms of both components will sensibly deviate from a sphere on account of their axial rotation and mutual tidal action, and the distortion will also cause the surface brightness to vary proportionally to the gravity. In such cases the formulae developed in the present paper would not be rigorously applicable, but still should offer a close approximation.<sup>6</sup> Effects of distortion upon the rotation factors could be taken into account by modifying the limits of integration and adopting an appropriate form for  $J(x, y, z)$ , but such a procedure is likely to involve a very heavy algebra.

Another consideration to be borne in mind when applying the present results to practical cases is the probable failure of the assumption underlying equation (3), that the star undergoing eclipse rotates like a rigid body. This is indeed not the case for the sun where the equatorial regions rotate by about 10 per cent faster than regions near the poles, and is still less likely to be true for rapidly rotating components of eclipsing binaries.<sup>7</sup> The equatorial velocity obtained with the aid of the present rotation factors would then be a certain mean of velocities derived from different layers which may either exceed somewhat, or be smaller than, the true velocity at the equator—depending on the geometry of each particular eclipse. On the other hand, provided that the law of increase of the angular velocity of rotation with diminishing latitude were known, it could be incorporated in the formulae for  $F$ , and very accurate observations might then enable us to obtain some information concerning the variation of angular velocity with latitude for stars other than the sun.

A further assumption underlying the present derivation of  $F$  is that the axis of rotation of the eclipsed star is constantly perpendicular to the orbital plane. In certain cases the validity of such an assumption may be problematic; for if the tides raised by the eclipsing star in the surface layers of its massive mate are small and if the latter are of minute density, no physical reason is known which should impel the massive star to keep its axis of rotation exactly perpendicular to the orbital plane. Moreover, a deviation from perpendicularity would entail no observational consequences that could in practice be detected otherwise than by the study of



the rotational effect. That is, if the axis of rotation is not perpendicular to the orbital plane, the rotational effect should no longer be symmetrical with respect to conjunction, and need not be zero at mid-primary minimum. The asymmetry of the rotational effect should also vanish and reappear periodically as the apparent inclination of the axis of rotation oscillates during the precessional motion. Such a situation may arise in systems where the component undergoing eclipse is very small or very massive compared with the eclipsing star. It seems indicated by observations, for instance, in the case of Algol;<sup>8</sup> but its discussion would widely exceed the scope of the present paper and is postponed for a future communication.

*Summary.*—Analytical formulae are set up for rigorous computation of the rotation factors required for the interpretation of the spectroscopically observed rotational effect in eclipsing binary systems. The formulae are exact as long as the components rotate like rigid bodies and their axes of rotation are perpendicular to the orbital plane. No notice will be taken at present of complications arising from the distortion of the components in form or from orbital eccentricity; and the components are assumed to appear as uniformly bright or limb-darkened circular discs. The rotation factors are found to be expressible in terms of the same functions that are required for an analysis of light changes of eclipsing variables. All such functions can be made to depend on two independent parameters—say, the ratio of radii  $k$  and the geometrical depth of the eclipse  $p$ —and tabulated in terms of these parameters for any particular phase of the eclipse. Tables already available for two such functions should greatly facilitate the computation of the rotation factors in practical cases.

<sup>1</sup> *Publ. D. A. O.*, 7, 133 (1938).

<sup>2</sup> *Proc. Am. Phil. Soc.*, in press.

<sup>3</sup> Provided that the rotation is direct. No case of retrograde rotation among eclipsing binaries has been detected so far.

<sup>4</sup> Analytical expressions for uniform discs have already been set up by Petrie (*op. cit.*). His formulae are, however, unnecessarily complicated. If we verify that, in Petrie's notations,

$$\sqrt{1 - y_1^2} + \sqrt{k^2 - y_1^2} = \delta,$$

his formulae become identical with those given presently.

<sup>5</sup> *Bull. Inst. Astr. Leningrad*, No. 45 (1939), and No. 50 (1940).

<sup>6</sup> The present formulae should then be in error by a factor of the order of  $r_g^1$ .

<sup>7</sup> Cf. Jeans, *Astronomy and Cosmogony*, Cambridge, 1928, Chapter X.

<sup>8</sup> Cf. McLaughlin, *Michigan Publ.*, 6, 3 (1934).

CERTAIN DIRECT PRODUCTS OF THE GROUPS OF  
SELF-ISOMETRIES

BY G. A. MILLER

Department of Mathematics, University of Illinois

Communicated March 2, 1942

The groups of rotations of the regular polygons are commonly called the dihedral groups. A necessary and sufficient condition that a dihedral group is non-abelian is that its order exceeds 4. If such a non-abelian group is the direct product of two of its proper subgroups one of these subgroups contains at most two invariant operators. If one of the two given factor groups contains two invariant operators the other contains no invariant operator besides the identity and hence it is a dihedral group whose order is twice an odd number. The former factor group is then of order 2. That is, *a necessary and sufficient condition that a non-abelian dihedral group is the direct product of two of its proper subgroups is that its order is divisible by 4 but not by a higher power of 2.* The only abelian dihedral group is the four group and hence it is the direct product of every two of its three proper subgroups.

The group of rotations of a regular polygon is also its group of self-isometries. That is, it is the most general group which transforms the regular polygon into itself in such a way that every pair of its points is transformed into a pair of points which are at the same distance from each other as the original points were. This results from the fact that such a group leaves the center of the polygon fixed and transforms its vertices transitively. If such a vertex is fixed another vertex can be transformed in two ways so that the order of the group of its self-isometries is twice the number of its vertices. If the number of the vertices is even a reflection of the polygon on its center corresponds to an invariant operator of order 2 in its group of self-isometries, but this group is the direct product of the subgroup generated by this invariant operator of order 2, which corresponds also to the rotation of the regular polygon in its plane through an angle of  $180^\circ$ , and some other proper subgroup of the group of self-isometries of the polygon if, and only if, the order of this group is divisible by 4 but by no higher power of 2, as was noted above.

The groups of movements of the five regular polyhedrons have received much attention and some of their abstract definitions were noted already by W. R. Hamilton (1805–1865); cf. *Bibliotheca Mathematica*, volume 10, page 314 (1910–1911). The groups of self-isometries of these polyhedrons have also been noted. For instance, on page 342 of volume 1 (1926) of his *Entwicklung der Mathematik*, Felix Klein refers to the group of self-isometries of the regular icosahedron, which is of order 120, and notes that it is

not the same as the symmetric group of degree 5 since it contains an invariant operator of order 2, but he does not point out here that it is the direct product of the group of movements of the regular icosahedron and a group of order 2. This fact is of fundamental importance because *all the groups of self-isometries of the five regular polyhedrons are the direct products of an invariant subgroup of order 2 contained therein and the corresponding group of movements of the regular solid*, and their abstract definitions can easily be obtained from the known abstract definitions of the groups of these solids.

If a group contains an invariant operator of order 2 and a subgroup of index 2 which does not include this invariant operator then it is the direct product of two proper subgroups of which one is the subgroup of order 2 generated by this invariant operator. From this obvious theorem it results directly that each of the groups of the self-isometries of the five regular solids is the direct product of two proper subgroups, one of these is the group of movements of the solid while the other is the group of order 2 generated by the operator of order 2 representing the reflections on the center of the regular solid, as was noted in the preceding paragraph. As the properties of such direct products are readily obtained from those of the given subgroup of index 2, it results that the groups of the self-isometries of the five regular solids present little which is of interest from the point of view of group theory beyond that of the groups of movements of these solids.

Since the group of movements of the cube is the same as the group of the regular octahedron and the group of the regular icosahedron is the same as the group of the regular duodecahedron there are three distinct groups of movements of the five regular solids. These are generated by two operators of orders 2 and 3 whose product has an order which is equal to one of the three numbers 3, 4, 5, respectively. From these relations it is easy to obtain abstract definitions of the three groups of self-isometries of the regular solids. In the case of the regular tetrahedron we may say that the group of its self-isometries is generated by two operators of orders 2 and 3, respectively, which satisfy the condition that the product is an operator of order 6 whose cube is invariant under the group. In fact, if  $s_1$  and  $s_2$  are these two generating operators of orders 2 and 3, respectively, we may observe that

$$1, s_1 i, s_2^2 s_1 s_2 i, s_1 s_2^2 s_1 s_2$$

is the four group and that this is invariant under the group generated by  $s_1$  and  $s_2$ .

Since  $s_1 i$  and  $s_2^2 s_1 s_2 i$  are both of order 2 they are commutative if their product is of this order. That this is the case results from the fact that  $(s_1 s_2)^3$  is equal to  $i$ . The given four group is invariant under  $s_1$  since  $s_1$  transforms each of the last two of the given operators into each other. It is

invariant under  $s_3$  because  $s_2$  transforms its three operators of order 2 cyclically. Hence the given four group and  $s_2$  generates the tetrahedral group which is invariant under  $i$  and the given relations constitute an abstract definition of the group of self-isometries of the regular tetrahedron. The quotient group of this group with respect to the invariant subgroup of order 2 is the tetrahedral group, and the operator of order 2 in the given defining relations had to be so chosen that it does not appear in the tetrahedral subgroup because the given operator of order 3 necessarily appears in this group.

A set of defining relations for the abstract group of order 48 which is simply isomorphic with the group of self-isometries of the cube is that its two generators of orders 2 and 4, respectively, have a product of order 6 whose cube  $i$  is invariant under this group of order 48. From  $s_1^2 = s_2^4 = (s_1s_2)^6 = 1$  and  $(s_1s_2)^3 = i$  it is easily seen that the group generated by  $s_1$  and  $s_2$  contains as an invariant subgroup the four group defined by

$$1, s_2^2, s_1s_2^2, s_2^2s_1$$

The second and third of these four operators are of order 2 by definition and the fact that the order of a transform of an operator by some other operator is the same as the order of the original operator. The order of the fourth of these operators is 2 as a result of the equation  $(s_1s_2)^3 = i$ . Hence the three given operators of order 2 are commutative and this four group is invariant under the entire group of order 48.

Since  $(s_1s_2)^2$  is of order 3 and transforms the three given operators of order 2 cyclically it and this four group generate the tetrahedral group. This tetrahedral group contains  $s_1s_2i$  as well as  $(s_1s_2)^2$ , and includes all the operators of odd order which appear in the given group of order 48. This contains two subgroups which are simply isomorphic with the symmetric group of order 24. One of these is obtained by extending the given tetrahedral group by  $s_1$  while the other is obtained by extending this subgroup by means of the operator  $s_1i$ . This group is simply isomorphic with the transitive permutation group of degree 6 and of order 48, which has received much attention in the theory of permutation groups in the early development of this subject. It represents the interchanges of the six plane surfaces of the square under its group of self-isometries.

A set of independent generators of the abstract group which is simply isomorphic with the group of self-isometries of the regular icosahedron can be selected in various ways. In the first place we may consider the abstract group which is generated by the two operators  $s_1$  and  $s_2$  of orders 2 and 3, respectively, which satisfy the condition that their product is of order 10 and that the fifth power of this product is an invariant operator of order 2. From  $(s_1s_2)^5 = i$  it results that  $s_1s_2i$  is an operator of order 5 and that  $s_1i, s_2$  are two operators of orders 2 and 3, respectively, whose product is of

order 5. It therefore results that they generate the icosahedral group as was noted already by W. R. Hamilton. Since  $i$  does not appear in this icosahedral group it follows that the group of isometries of the regular icosahedron is simply isomorphic with the direct product of the icosahedron group and the group of order 2, as was noted above. An abstract definition of this direct product is similarly seen to be the group generated by an operator of order 2 and an operator of order 5 whose product is of order 6, but has a cube which is commutative with these two operators.

## CERTAIN CONGRUENCE CRITERIA CONNECTED WITH FERMAT'S LAST THEOREM

BY H. S. VANDIVER

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF TEXAS

Communicated February 16, 1942

In another paper<sup>1</sup> the writer obtained a result which may be stated as follows:

THEOREM I. *If  $l$  is an odd prime and*

$$x^l + y^l + z^l = 0 \tag{1}$$

*and  $y \equiv 0 \pmod{l}$ ;  $(xz, l) = 1$ , the prime ideal  $\mathfrak{p}$  in  $k(\zeta)$ , with  $\zeta = e^{2i\pi/l}$  is a divisor of the rational prime  $p$ ;*

$$c = \frac{N(\mathfrak{p}) - 1}{l}$$

*where  $N(\mathfrak{p})$  is the norm of  $\mathfrak{p}$ , further,  $\mathfrak{p}$  belongs to an exponent<sup>2</sup> which is prime to  $l$  with  $xyz \not\equiv 0 \pmod{p}$ , then*

$$\binom{c}{s} + \binom{c}{s+l} t^l + \binom{c}{s+2l} t^{2l} + \dots \tag{2}$$

*is divisible by  $p$  for  $s = 1, 2, \dots, l-1$ , and  $y/x = t$ .*

The present article will be devoted to extensions of the methods which were employed in the proof of this theorem. A related method is employed in the proof required for the demonstration of a theorem in another article.<sup>3</sup> A companion formula to (2) which is obtained by the method<sup>1</sup> we previously referred to is

$$\frac{1}{x^c} \equiv \binom{c}{0} + \binom{c}{l} t^l + \binom{c}{2l} t^{2l} + \dots \tag{3}$$

The condition  $xyz \not\equiv 0 \pmod{p}$  is of course, satisfied if we take  $p$  sufficiently large in (2) and (3). Also, we note that in view of (1) we can employ  $z$  in place of  $x$  in our argument, hence (2) holds for  $t = y/x$  and  $t_1 = y/z$ . For  $p$  sufficiently large these are incongruent modulo  $p$ . Hence (2) holds for two incongruent values of  $t$ .

Applying our second method<sup>3</sup> we obtain from (1), under the assumption that the second factor of the class number of the field defined by  $\zeta$  is prime to  $l$ , the congruence

$$(x + \zeta^a y)^c \equiv (x + \zeta^{-a} y)^c \pmod{p} \quad a = 0, 1, 2, \dots, l-1 \quad (4)$$

This gives<sup>4</sup>

$$\begin{aligned} & \binom{c}{s} x^s y^{c-s} + \binom{c}{s+l} x^{s+l} y^{c-(s+l)} + \dots \\ & \equiv \binom{c}{l-s} x^{l-s} y^{c-(l-s)} + \binom{c}{2l-s} x^{2l-s} y^{c-(2l-s)} + \dots \end{aligned} \quad (5)$$

and we may state the

**THEOREM II.** *If (1) is satisfied under the conditions given in Theorem I and the second factor of the class number of  $k(\zeta)$  is prime to  $l$ , then for each prime  $p$  exceeding a certain limit the congruences (5) hold, and they also hold when  $x$  is replaced by  $z$ ;  $xyz \not\equiv 0$ .*

Elsewhere<sup>3, 4</sup> the congruences (5) were employed in a bit different way. There,  $p$  was in the particular form  $1 + cl$  and we considered smaller values for  $c$ . Employing this idea we were able to obtain criteria which led to the proof of Fermat's last Theorem for all prime exponents  $< 619$ . We also find<sup>5</sup> from (1)

$$(x + \zeta^k z)^c \equiv (\zeta^k x + z)^c \pmod{p} \quad (6)$$

$$k = 0, 1, 2, \dots, l-1$$

which gives

$$\begin{aligned} & \binom{c}{s} x^s y^{c-s} + \binom{c}{s+l} x^{s+l} y^{c-(s+l)} + \dots \\ & \equiv \binom{c}{s} z^s x^{c-s} + \binom{c}{s+l} z^{s+l} x^{c-(s+l)} + \dots \end{aligned} \quad (7)$$

which provides another criterion for the solution of (1) in addition to Theorem II.

Much work has been done on Fermat's Last Theorem in connection with the congruence

$$x^l + y^l + z^l \equiv 0 \pmod{p} \quad (8)$$

where  $p \neq l$ . It is known<sup>6</sup> that if  $p$  exceeds a certain limit this congruence always has solutions with  $x, y$  and  $z$ , prime to  $p$ . This may also be true in connection with the congruences (5) and (7) but it does not appear obvious why this is so. (5) is obviously satisfied when  $y \equiv 0(\text{mod } p)$  but  $p$  can be taken  $> y$ . Similarly (7) is satisfied if either  $x + z$ , or  $x - z$  is divisible by  $p$  but again  $p$  can be taken  $> x + z$  or  $x - z$ . For the smaller values of  $c$ , the only solution of (5) is  $y \equiv 0(\text{mod } p)$ .

In the statement of Theorem I,  $p$  was limited to primes belonging to an exponent which was prime to  $l$ . Now the writer has shown<sup>7</sup> that for there to exist an ideal not of the first degree in the basis classes  $C$  of the irregular class group of  $k(\zeta)$  defined by

$$C^s = r^{l-2a} = 1, \quad (8a)$$

$r$  being a primitive root of  $l$ , it is necessary that  $l - 2a$  and  $l - 1$  have a common factor,  $> 1$ . It will be found that this condition is also necessary when we take classes  $C_1$  of the irregular class group of  $k(\zeta)$  defined by

$$C_1^s = r^{3a} = 1.$$

If we apply this necessary condition to all the fields  $k(\zeta)$  for  $< 619$  we find that there are only eight values for  $l$  in this range where it is possible for these types of classes to contain other than ideals of the first degree. If it is true that all of these irregular classes contain only ideals of the first degree, then Theorem I may be greatly simplified, as we need only state that  $p$  in (2) belongs to some exponent  $> 1$  modulo  $l$ .

From the congruences (2) we may now obtain congruences independent of  $t$ . For, we may take (2) for a particular value of  $s$  and multiply this congruence by

$$t^l, t^{2l}, \dots, t^{l(c-1)}; \quad (9)$$

in this way we obtain a set of  $c$  congruences, which yield the relation

$$D \equiv 0(\text{mod } p), \quad (10)$$

where  $D$  is the circulant

$$\begin{vmatrix} a_0 & a_1 & a_2 & \dots & 0 \\ 0 & a_0 & a_1 & a_2 & \dots & 0 \\ 0 & 0 & a_0 & a_1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_1 & a_2 & a_3 & a_4 & \dots & a_0 \end{vmatrix} \quad (11)$$

and where, if  $i = 1, 2, \dots, h$ ,

$$a_i = \binom{c}{s + il}$$

$$h = \left[ \frac{c}{l} \right],$$

$[u]$  being the greatest integer in  $u$ . Now we may obtain a simpler determinant when  $p^f \equiv 1 \pmod{l^2}$  and  $f > 1$  with  $c_1 l \equiv c$ , for if we multiply the congruences (2) by

$$t^1, t^2, \dots, t^{l(c_1 - 1)}$$

we obtain a set of congruences from which we may infer, if we note that since  $f > 1$ ,  $t^c \equiv 1 \pmod{p}$ ;  $D_1 \equiv 0 \pmod{p}$ , where

$$D_1 = \begin{vmatrix} a_0 & a_1 & \dots & a_{c_1 - 1} \\ a_{c_1 - 1} & a_0 & \dots & a_{c_1 - 2} \\ \dots & \dots & \dots & \dots \\ a_1 & a_2 & \dots & a_0 \end{vmatrix} \quad (11a)$$

for  $s > 0$ .

If we now take (2) again for a fixed  $s$  and multiply this congruence by each of the quantities in (10), and consider any  $(c - 1)$  of the resulting congruences as a set of linear congruences in the quantities in (9), and employing the method used in deriving Cramer's rule, we find that

$$\Delta t^1 \equiv A \pmod{p} \quad (12)$$

where  $\Delta$  is now the determinant formed by the coefficients of the quantities in (9), and where  $(A)$  is obtained from the determinant  $\Delta$  by replacing the elements in the first column by the elements in the congruences which do not involve  $t$ . Now (12) is a congruence in  $t$  which is satisfied by each of the quantities

$$t, \zeta t, \zeta^2 t, \dots, \zeta^{l-1} t.$$

Now consider the modulus  $\mathfrak{p}$ , a prime ideal divisor of  $p$ , and take in addition  $t_1$  in place of  $t$ , in (12). The relation  $t_1 \equiv t \zeta^b \pmod{p}$  gives  $t^1 - t_1^1 \equiv 0 \pmod{p}$ , and this is impossible for  $p$  sufficiently large. Then, (12) with  $\mathfrak{p}$  in place of  $p$  has more than  $l$  incongruent solutions unless  $\Delta \equiv 0 \pmod{p}$  in which case  $A \equiv 0 \pmod{p}$ . If we also consider the relation

$$\Delta t^{kl} \equiv A_k \pmod{p}$$

obtained from our former congruences, it will therefore be seen that

$$A_k \equiv 0 \pmod{p}.$$



By taking  $s = 1, 2, \dots, l - 1$ , we obtain in this way a number of determinants each divisible by  $p$ . If  $p^f \equiv 1 \pmod{l^2}$  we obtain as in the derivation of (11a) a set of simpler determinant criteria.

We now apply methods similar to those used in the derivation of Theorem I to the first case of Fermat's Last Theorem. In (1) we now assume that

$$(xyz, l) = 1$$

and we select a prime ideal  $\mathfrak{p}$  in  $k(\zeta)$  such that if  $\lambda = (1 - \zeta)$ ,

$$\mathfrak{p} = (1 + m\lambda^l + \theta\lambda^{l+1}) = (\omega) \quad (13)$$

where  $m$  is a rational integer  $\not\equiv 0 \pmod{l}$ . By a known theorem<sup>8</sup> we may find an infinity of  $\theta$ 's, if  $m$  is fixed, so that an expression of this form gives an infinity of prime ideals. Since  $\omega$  is primary, we have<sup>9</sup>

$$\left(\frac{x + y\zeta^s}{\omega}\right) = \left(\frac{\omega}{x + y\zeta^s}\right) = 1 \quad (14)$$

The relation (13) shows that  $\mathfrak{p}$  is an ideal of the first degree, hence

$$p = 1 + cl.$$

Then we may write, using (14), and if  $xy \not\equiv 0 \pmod{p}$ ,

$$(x + y\zeta^s)^c \equiv 1 \pmod{\mathfrak{p}} \quad (15)$$

where  $s = 0, 1, \dots, l - 1$ , if we assume that  $x + \zeta^s y \not\equiv 0 \pmod{p}$ . On expanding the left hand member of this congruence and using the various values of  $s$  we obtain congruences of the same type as (2) for  $s = 1, 2, \dots, l - 1$ , and similarly we find that a determinant of the same type as (11) is divisible by  $p$ . In the discussion of the second case of our theorem, we took values of  $p$  large enough so that various congruences involving  $x$ ,  $y$  and  $z$  did not hold. However, for the first case, which we are now discussing, this is not necessary. For, if  $x + \zeta^a y \equiv 0 \pmod{\mathfrak{p}}$ , then since  $\not\equiv 0 \pmod{p}$ ,  $x + \zeta^s y \not\equiv 0 \pmod{p}$  for  $s \not\equiv a \pmod{p}$ . Under this condition (5) gives

$$x^c(\zeta^a - \zeta^s)^c \equiv 1 \pmod{\mathfrak{p}} \quad (16)$$

and (1) gives

$$x + z = v^l$$

yielding

$$x \equiv v^l \pmod{p}$$

or

$$x^c \equiv 1 \pmod{p}$$

hence from (16) we have

$$(\zeta^a - \zeta^s)^c \equiv 1 \pmod{p}$$

but by a known result<sup>10</sup> this is impossible since  $m \not\equiv 0 \pmod{l}$  in (13).

We may then state the

**THEOREM III.** *Let  $l$  be an odd prime,*

$$x^l + y^l + z^l = 0$$

*and  $y \equiv 0 \pmod{l}$ ;  $(xz, l) = 1$ ; the prime ideal  $\mathfrak{p}$  is a divisor, in the field  $k(\zeta)$ ;  $\zeta = e^{2\pi i/l}$ , of the rational prime  $p$ ;*

$$c = \frac{N(\mathfrak{p}) - 1}{l}$$

*where  $N(\mathfrak{p})$  is the norm of  $\mathfrak{p}$ . Further,  $\mathfrak{p}$  belongs to an exponent which is prime to  $l$ ; then for any  $p$  sufficiently large the determinant (11) is divisible by  $p$ , with  $i = 1, 2, \dots, h$ ,*

$$a_i = \begin{pmatrix} c \\ s + il \end{pmatrix}$$

$$h = \left[ \frac{c}{l} \right],$$

*for  $s = 1, 2, \dots, l - 1$ . Also  $[u]$  is the greatest integer in  $u$ . In addition, certain minors of (11), as indicated in the relation (12), are divisible by  $p$ . In particular, if  $p$  belongs to an exponent  $> 1$ , and  $p^f \equiv 1 \pmod{l^2}$  the determinant (11a) is divisible by  $p$  for each  $s$  mentioned above.*

We also have

**THEOREM IV.** *Let  $l$  be an odd prime,*

$$x^l + y^l + z^l = 0$$

*with  $(xyz, l) = 1$ , and  $\mathfrak{p}$  is a prime ideal in  $k(\zeta)$  such that*

$$\mathfrak{p} = (1 + m\lambda^l + \theta\lambda^{l+1})$$

*where  $m \not\equiv 0 \pmod{l}$ . Further set  $N(\mathfrak{p}) = p$  with  $c$  defined as in Theorem II. Then the determinant (11) is divisible by  $p$ , for  $s = 1, 2, \dots, l - 1$ ,  $a_i$  and  $h$  being defined as in Theorem II. Also certain minors of (11) as indicated in the relations (12) are divisible by  $p$ .*

The statements of Theorems I and III each include the condition " $\mathfrak{p}$  belongs to an exponent which is prime to  $l$ ;" we have, however, as already indicated no convenient criteria for determining if a given prime ideal satisfies such a condition. We shall now point out, however, that there are certain classes of ideals which may be immediately recognized as belonging

to an exponent prime to  $l$ . For if  $p$  belongs to the exponent  $f(\text{mod } l)$  and such that  $ef = l - 1$  with  $e$  odd, and if  $r$  is a primitive root of  $l$ ,

$$p(\zeta) = p(\zeta^{r^e})$$

which gives from (8a)

$$r^{e(l-1)} \equiv 1 (\text{mod } l)$$

which is impossible since  $r$  is a primitive root of  $l$  and the exponent is odd.

<sup>1</sup> *Proc. Nat. Acad. Sci.*, 15, 45 (1929).

<sup>2</sup> We say that an ideal  $\mathfrak{p}$  belongs to an exponent  $n$  when  $\mathfrak{p}^n$  is a principle ideal but  $\mathfrak{p}^d$  is not a principal ideal for  $0 < d < n$ .

<sup>3</sup> *Trans. Amer. Math. Soc.*, 31, 632 (1929).

<sup>4</sup> *Duke Math. Jour.*, 5, 419-420 (1939).

<sup>5</sup> *Trans. Amer. Math. Soc.*, 31, 632-633 (1929).

<sup>6</sup> Dickson, *Jour. für. Math.*, 135, 134-141 (1909).

<sup>7</sup> *Proc. Nat. Acad. Sci.*, 16, 303 (1930).

<sup>8</sup> Landau, *Math. Zeitschrift*, 2, 52-154 (1918).

<sup>9</sup> Hasse, *Jahresbericht Deutschen Math.-Verein*, Ergänzungsbände VI, 61 (1930).

<sup>10</sup> Hasse, loc. cit., p. 112.

## ON THE PHYSICAL CHARACTERISTICS OF THE HYDRA CLUSTER OF NEBULAE

BY F. ZWICKY

NORMAN BRIDGE LABORATORY OF PHYSICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated March 12, 1942

*A. Observational Data.*—In a previous paper<sup>1</sup> counts of the brighter nebulae in the Hydra cluster were communicated. The distance of the Hydra cluster was estimated as  $7.3 \times 10^6$  parsecs and the average apparent velocity of recession to be expected was given as  $v = 4100$  km./sec. No observations of the red shift for nebulae in the Hydra cluster were available at the time. Dr. Hubble subsequently obtained a spectrum plate for NGC 3309 which nebula is a member of the Hydra cluster. He kindly informs me that the apparent velocity of recession for NGC 3309 is of the order of 3950 km./sec., a value which is in good agreement with our original estimate.

*B. The Radial Distribution of Nebulae in the Hydra Cluster.*—As emphasized previously<sup>1</sup> the Hydra cluster exhibits spherical symmetry and is therefore suspected to have reached a statistically stationary state. In order to check this conclusion two additional tests are available. In the

first place we may compare the observed radial distribution of nebulae in the Hydra cluster with the distribution derived by Emden<sup>2</sup> for the bounded isothermal gravitational gas sphere and secondly we can verify whether the ratio  $\overline{w_r^2}/\rho_0$  of the square of the dispersion in radial velocities of the cluster nebulae to the central density  $\rho_0$  of the cluster can be correctly determined from the observed *structural length or structural index*  $\alpha$  of the cluster.<sup>3</sup>

If from the previously<sup>1</sup> given counts in the Hydra cluster of nebulae brighter than the apparent photographic magnitude  $m_p = 16.2$  we subtract the nebulae which belong to the general background (1.8 nebulae per square degree), we obtain for the average number of nebulae  $N_r$  per square degree at various distances  $r$  from the center of the cluster the values given in table 1.

TABLE I  
RADIAL DISTRIBUTION OF NEBULAE IN THE HYDRA CLUSTER

$r$ IN MINUTES OF ARC	$N_r$	$8.2 N_r$	$r_1$	$1000 D$	$1000 D - 37$
0	...	....	0	3032	2995
10	241.8	1983	2.5	1815	1778
20	121.2	994	5	832	795
30	44.6	366	7.5	476	439
40	24.9	204	10	318	281
50	30.7	252	12.5	243	206
60	24.9	204	15	195	158
70	13.3	109	17.5	163	126
80	11.6	95	20	140	103
90	7.7	63	22.5	126	89
100	1.9	15.6	25	111	74
110	6.4	52.5	27.5	97	60
120	5.3	43.5	30	92	55
130	1.0	8.2	32.5	87	50
140	4.2	34.4	35	81	44
150	5.4	44.3	37.5	75	38
160	...	....	40	69.6	32.6
170	2.4	19.7	42.5	66.4	29.4
180	2.1	17.2	45	63.1	26.1
190	2.6	21.3	47.5	59.8	22.8
200	2.4	19.7	50	56.5	19.5
210	1.3	9.1	52.5	54.6	17.6
220	0.6	4.9	55	52.7	15.7
230	2.6	21.3	57.5	50.8	13.8
240	1.3	9.1	60	49	12
250	0.5	4.6	62.5	45	8
280	...	....	70	41.5	4.5
340	...	....	85	37	0

Column 4 gives the distances  $r_1$  in the standard<sup>3</sup> Emden isothermal gas sphere to which the actual distances  $r$  from the center of the cluster are

reduced through the relation  $r = \alpha r_1$ . The values of  $D$  represent the spatial densities in an infinite isothermal sphere while  $D - 37/1000$  is the projected density in that bounded isothermal sphere which best

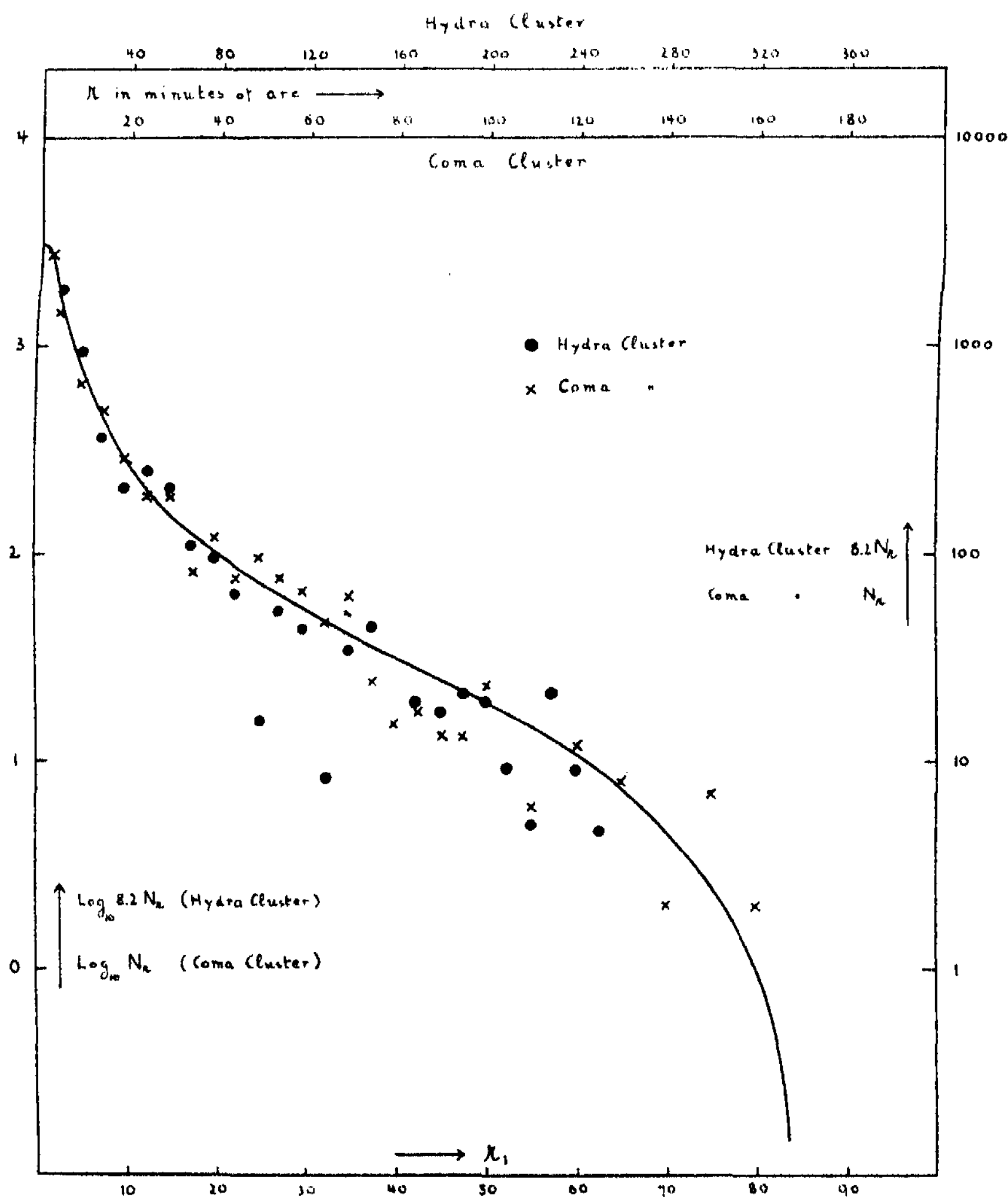


FIGURE 1

represents the radial distribution of the brighter nebulae in the Coma cluster with which we here compare the data on the Hydra cluster. From figure 1 we see that the properly reduced radial distributions of the Coma cluster and of the Hydra cluster agree among themselves, as well as with

the theoretical curve, within the limits of statistical fluctuations to be expected for the relatively small number of nebulae involved in the counts.

There is some indication that the values of  $N_r$  for the Hydra cluster begin to fall off from the theoretical curve at slightly smaller values of  $r_1$  than those for the Coma cluster. This checks the general expectation that, the smaller the population of a cluster (and the higher its velocity dispersion) the smaller should be the distance from the center of the cluster at which the Boltzmann distribution changes over into a Smoluchowski distribution.

The data for the Coma cluster plotted in figure 1 include the nebulae brighter than the apparent magnitude  $m_p = 16.6$  or brighter than the absolute magnitude  $M_p = -14.5$  while the data for the Hydra cluster include the nebulae brighter than the apparent magnitude  $m_p = 16.2$  or the absolute magnitude  $M_p = -13.1$ .

*C. Physical Characteristics of the Hydra Cluster.*—From figure 1 we deduce that the structural length  $\alpha = r/r_1$  for the Coma and Hydra clusters in angular measure are  $2'$  arc and  $4'$  arc, respectively. Since the distances of the two clusters are  $13.8 \times 10^6$  and  $7.3 \times 10^6$  parsecs, respectively, we obtain for the

$$\text{Coma Cluster:} \quad \alpha = 2.48 \times 10^{22} \text{ cm.} \quad (1)$$

$$\text{Hydra Cluster:} \quad \alpha = 2.56 \times 10^{22} \text{ cm.} \quad (2)$$

By definition<sup>3</sup> it is

$$\alpha = (\overline{w^2}/12\pi\Gamma\rho_0)^{1/2} \quad (3)$$

where  $\overline{w^2}$  is the square of the velocity dispersion in a stationary cluster and  $\rho_0$  is its central density. Since the velocity dispersion for the nebulae in the Hydra cluster is not known at the present we give in table 2 values for  $\rho_0$  in dependence of a number of values for the velocity dispersion.

TABLE 2

$(\overline{w^2})^{1/2}$ IN KM./SEC.	$\rho_0$ IN G./CM. <sup>3</sup>	$\overline{m}/m_\odot$
250	$3.8 \times 10^{-25}$	$5.8 \times 10^9$
500	$1.5 \times 10^{-24}$	$2.3 \times 10^{10}$
750	$3.4 \times 10^{-24}$	$5.2 \times 10^{10}$
1000	$6.0 \times 10^{-24}$	$9.2 \times 10^{10}$
1500	$1.4 \times 10^{-23}$	$2.1 \times 10^{11}$
2000	$2.4 \times 10^{-23}$	$3.7 \times 10^{11}$
2500	$3.8 \times 10^{-23}$	$5.8 \times 10^{11}$

In column 3 are given the ratios of the average mass  $\overline{m}$  of the nebulae involved to the mass  $m_\odot$  of the sun which correspond to the different values of the velocity dispersion given in column 1. The method of calculation

of the ratios  $\bar{n}/\bar{n}_\odot$  is given in another place<sup>8</sup> where it is shown that for the same values of  $(\bar{w}^2)^{1/2}$  the ratios  $\bar{n}/\bar{n}_\odot$  for the Coma cluster are about half as large as those here obtained for the Hydra cluster. Now the nebulae included in the counts of the Coma cluster are brighter than the absolute magnitude  $M_p = -14.5$  while those for the Hydra cluster are brighter than  $M_p = -13.1$ . Therefore if the same types of nebulae are included in the two clusters  $\bar{n}$  (Coma) should be slightly larger than  $\bar{n}$  (Hydra). This means that  $2\bar{w}^2$  (Hydra) should be slightly smaller than  $\bar{w}^2$  (Coma) or the velocity dispersion in the Hydra cluster should be smaller than that in the Coma cluster by a factor slightly smaller than 0.7. It will be interesting to check this conclusion by an investigation of the velocity dispersion in the Hydra cluster.

From our data we can also answer the question regarding the central density, that is the number of nebulae, say, per cubic mega-parsec ( $10^{18}$  cubic parsecs) in the two clusters. It was shown in another place<sup>8</sup> that

$$\rho_0 = \sigma_0/3.03\alpha \quad (4)$$

where  $\rho_0$  is either the central spatial density or the number of nebulae per unit cube while  $\sigma_0$  is the projected central density. From the extrapolation of our counts of nebulae to the center of the two clusters we consequently obtain the following results.

In the center of the Hydra cluster we obtain per cubic mega-parsec  $8.7 \times 10^5$  nebulae which are brighter than the absolute magnitude  $M = -13.1$  while the corresponding number for the Coma cluster is  $2 \times 10^6$  nebulae per cubic mega-parsec which are brighter than the absolute magnitude  $M = -14.5$ . In comparison we mention that in the general field the average number of nebulae in the same range of absolute magnitudes is less than ten per cubic mega-parsec. The number of nebulae per unit volume in the center of the large clusters is therefore considerably higher than even experts in nebular counts might have expected. The high nebular density in clusters perhaps sheds some light on the remarkable fact that large symmetrical clusters of nebulae show all of the physical characteristics required for statistically stationary assemblies of objects whose interactions are governed by Newton's law. The peculiar possibility that these giant clusters whose elementary building stones are the nebulae themselves should provide the first quantitative test for Emden's results which were derived for gaseous spheres whose elementary building stones are atoms or molecules requires of course additional and thorough investigation.

In conclusion we mention that the clusters in Perseus, Cancer, Fornax and Pegasus as well as some of the groups in Pisces lend themselves to the

same analysis as that applied here to the Hydra cluster and in another place to the Coma cluster.<sup>3</sup>

<sup>1</sup> Zwicky, F., these PROCEEDINGS, 27, 264 (1941).

<sup>2</sup> Emden, R., *Gaskugeln*, Teubner, Leipzig, 1907.

<sup>3</sup> Zwicky, F., *Th. von Kármán Anniversary Volume*, May, 1941. See also a paper on the clustering of nebulae which is to appear shortly in the *Astrophys. Journal*.





# PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

Volume 28

May 15, 1942

Number 5

Copyright 1942 by the National Academy of Sciences

## *RADIOACTIVE CARBON AS AN INDICATOR OF CARBON DIOXIDE UTILIZATION. IX. THE ASSIMILATION OF CARBON DIOXIDE BY PROTOZOA*

BY C. B. VAN NIEL, J. O. THOMAS, S. RUBEN AND M. D. KAMEN

THE HOPKINS MARINE STATION (OF STANFORD UNIVERSITY), PACIFIC GROVE, CALIFORNIA, THE CHEMICAL LABORATORY, AND THE RADIATION LABORATORY OF THE UNIVERSITY OF CALIFORNIA

Communicated March 23, 1942

Previous publications from these laboratories have demonstrated the widespread occurrence of carbon dioxide utilization by living cells.<sup>1-7</sup> A survey of the known cases<sup>8</sup> has made it extremely likely that in all cases the absorption leads primarily to the formation of a carboxylic acid in which the carbon dioxide molecule becomes incorporated as a carboxyl group.

So far, the class of protozoa had not yet been investigated in this respect. It is, however, noteworthy that Rahn<sup>9</sup> has recently published evidence for the absolute necessity of CO<sub>2</sub> in the development of these organisms, a fact which is entirely in line with the general conclusions regarding the rôle of this compound in metabolism, expressed in the previous paper of this series.<sup>8</sup>

Experiments by one of us (J. O. T.) with pure cultures of the holotrichous ciliate, *Tetrahymena geleii*, have shown that under anaerobic conditions a sugar fermentation occurs in which the main metabolic products are lactic, acetic and succinic acids, accompanied by carbon dioxide. Furthermore, it was ascertained that suspensions of the protozoan in a phosphate-bicarbonate buffer can take up measurable amounts of carbon dioxide. In view of the work of Wood and Werkman,<sup>10-12</sup> Carson and Ruben<sup>2, 3</sup> and others, it seemed probable that this assimilation would be responsible for the production of succinic acid. Since it is difficult, if not impossible, due to the net evolution of CO<sub>2</sub> to demonstrate conclusively the participation of carbon dioxide in the succinic acid formation we have studied the anaerobic metabolism of *Tetrahymena geleii* in the presence of radioactive C<sup>11</sup>O<sub>2</sub>.\*

The ciliate was grown in yeast extract media with 2% glucose under conditions of adequate aeration. The cells were centrifuged for use in the experiments, washed and resuspended in phosphate buffer (pH 7.5) with 1%

glucose. The suspension was then shaken in the presence of  $C^{11}O_2$  in an oxygen-free atmosphere for 30 minutes at  $30^\circ C$ . Following the addition of small amounts of lactic, acetic, pyruvic, succinic and fumaric acids as carriers the suspension was rapidly boiled in order to stop further metabolism, and to drive off dissolved  $CO_2$ . Complete elimination of residual  $C^{11}O_2$  was, as usual, accomplished by addition of  $NaHCO_3$ , acidification and further boiling. Hereafter the suspensions were centrifuged, the cells resuspended in  $H_2O$  and again thrown down. Measurements of the insoluble cellular material at this time showed it to be practically without radioactivity.

Pyruvic and fumaric acids were precipitated from separate aliquots of the original supernatant solution with 2,4-dinitrophenylhydrazine and mercurous nitrate, respectively. These precipitates, also, showed but little activity.

From another aliquot of the original supernatant the volatile acids were removed by a vacuum distillation, and the activities of the distillate and the residue measured separately. The latter fraction contains the known products, succinic and fumaric acids.

The relative activities of the various fractions are assembled in table 1.

TABLE 1  
RELATIVE ACTIVITIES OF THE FRACTIONS OBTAINED FROM A SUSPENSION OF *Tetrahymena geleii*, INCUBATED FOR 30 MINUTES IN THE PRESENCE OF  $C^{11}O_2$

FRACTION	ACTIVITY IN % OF TOTAL $C^{11}O_2$ REDUCED <sup>a</sup>
Fumaric acid	< 4
Pyruvic acid	< 0.2
Volatile acid	< 0.1
Non-volatile residue	98
Cell material	2

<sup>a</sup> All values corrected for decay, hence strictly comparable.

The uptake of  $C^{11}O_2$  was surprisingly large; during the incubation period about 35% of the available supply had been reduced by the protozoa. It even reached 50% in a subsequent experiment. It is important to note that the two principal products of the fermentation, lactic and acetic acid, were completely inactive, and only a small fraction of the activity was contained in the cell material after one washing with  $H_2O$ . The non-volatile residue, containing the succinic and fumaric acids, accounted for practically all the activity. Boiling with acid permanganate reduced this activity only slightly. It may be pointed out that the activity of the fumaric acid fraction may have been caused partly by co-precipitated mercurous succinate.

These results make it probable, therefore, that the  $C^{11}O_2$  had actually been used in the production of succinic acid. For a convincing demon-

stration a characterization of the nature of the radioactive substance was carried out in an additional experiment in the following manner.

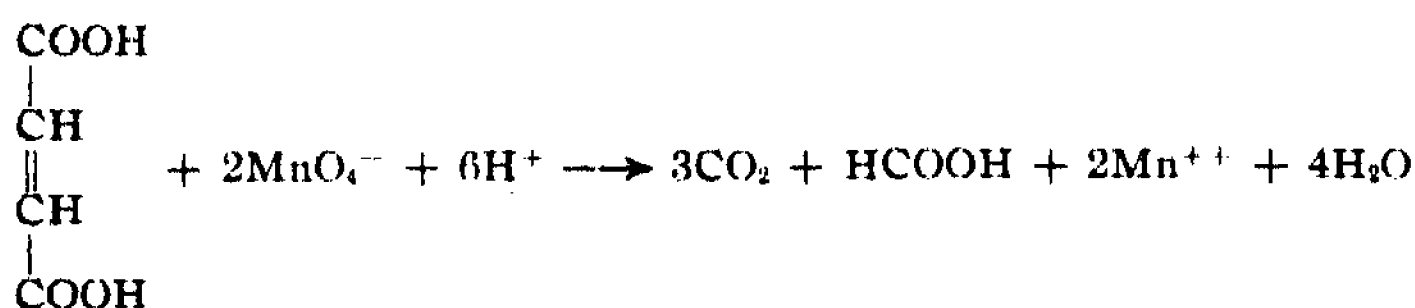
The protozoa suspension, prepared in the same manner as before, and after having been shaken in the presence of  $C^{11}O_2$  for 30 minutes at  $30^\circ C.$ , was again acidified, boiled and centrifuged. The supernatant, upon the addition of about 20 mg. carrier succinic acid, was neutralized (final pH 7.17), and subjected to the action of a succino-dehydrogenase preparation from beef heart in the presence of air.<sup>†</sup> The dehydrogenation of the succinic acid was followed manometrically, and was complete in 30 minutes at  $37^\circ C.$  The enzyme preparation was then precipitated by the addition of trichloroacetic acid, and the precipitate removed by centrifugation. The supernatant, which now contained the original succinic acid in the form of fumaric acid, was divided into three aliquots which were used for the following tests:

1. The total radioactivity was measured in one aliquot directly. It showed that 50% of the available  $C^{11}O_2$  had been assimilated.

2. In the second aliquot the fumaric acid was precipitated with mercurous nitrate, the precipitate filtered, washed, dried and its activity determined.

3. The third fraction was treated with  $KMnO_4$  in 1.5 *N*  $H_2SO_4$  at  $40^\circ C.$  in a stream of nitrogen. The gas evolved was passed through  $Ba(OH)_2$  solution, in which a copious precipitate of  $BaCO_3$  occurred. This also was filtered, washed, dried and weighed, and its activity was measured on weighed portions. With the solution remaining after the permanganate treatment a steam distillation was performed, and the distillate, after having been made alkaline, measured in the Geiger counter.

The oxidation of fumaric acid under the conditions of the test proceeds in accord with the equation



It has been demonstrated (Allen and Ruben<sup>15</sup>) that the formic acid arises exclusively from the methine carbons of the fumaric acid. Thus, a comparison of the activities of the oxidation products,  $CO_2$  and formic acid, permits of an unambiguous localization of the  $C^{11}$  atoms in the fumaric, and, therefore, in the original succinic acid. Table 2 presents the results.

From the first experiment (table 1) it has emerged that the amount of radioactive fumaric acid in the fermentation solution is certainly not above 4%. The very large quantity of this substance in a solution subjected to

the highly specific action of succino-dehydrogenase thus shows conclusively that at least 83% of the  $C^{11}O_2$  reduced had been converted by *Tetrahymena geleii* into succinic acid. This value must be considered as a lower limit because unfortunately the quantity of mercurous fumarate was large, so that considerable self-absorption of the  $C^{11}$  positrons seriously interfered with an accurate radioactivity measurement. This is borne out by the fact that the determination of  $C^{11}$  in the  $BaCO_3$ , carried out with a smaller amount (thinner sample), yielded a higher relative activity, undoubtedly the result of a decreased self-absorption.

A comparison of the values for the carbonate and formic acid fractions clearly proves that the labeled carbon was present in the fumaric acid only in the carboxyl groups. Together with the evidence given above, this implies that the carbon dioxide assimilated during the anaerobic sugar metabolism of *Tetrahymena geleii* is converted mainly into succinic acid, and is present exclusively in the carboxyl groups of this substance.

TABLE 2  
RELATIVE ACTIVITIES OF THE VARIOUS FRACTIONS AND ALIQUOTS

FRACTION	ACTIVITY IN % OF TOTAL $C^{11}O_2$ REDUCED
Fumaric acid as mercurous fumarate	83
$CO_2$ produced during $KMnO_4$ oxidation of second aliquot	89
Formic acid fraction of fumarate oxidation	1

Since the mechanism for the formation of succinic acid with the above properties has been adequately dealt with elsewhere<sup>8, 12, 13</sup> it need not be further discussed at this place.

We are indebted to Professor E. O. Lawrence and members of the Radiation Laboratory for making these experiments possible.

*Summary.*—Experiments with radioactive carbon dioxide have shown that the formation of succinic acid during the anaerobic sugar metabolism of the holotrichous ciliate, *Tetrahymena geleii*, involves the assimilation of carbon dioxide. Since the radioactivity of the succinic acid so produced is strictly limited to the carboxyl groups, the results support the current concepts of the mechanism of succinic acid formation.

\* The half-life of  $C^{11}$  is 20.5 minutes; the production and quantitative determination of  $C^{11}$  have been described elsewhere.<sup>14</sup>

† We are indebted to Mr. Bartley Carden for the enzyme preparation and for directions concerning its use.

<sup>1</sup> Ruben, S., and Kamen, M. D., *Proc. Nat. Acad. Sci.*, **26**, 418 (1940).

<sup>2</sup> Carson, S. F., Ruben, S., Kamen, M. D., and Foster, J. W., *Ibid.*, **27**, 475 (1941).

<sup>3</sup> Carson, S. F., and Ruben, S., *Ibid.*, **26**, 422 (1940).

<sup>4</sup> Barker, H. A., Ruben, S., and Kamen, M. D., *Ibid.*, **26**, 426 (1940).

<sup>5</sup> Barker, H. A., Ruben, S., and Beck, J. V., *Ibid.*, **26**, 477 (1940).

<sup>6</sup> Carson, S. F., Foster, J. W., Ruben, S., and Barker, H. A., *Ibid.*, **27**, 229 (1941).

<sup>7</sup> Foster, J. W., Carson, S. F., Ruben, S., and Kamen, M. D., *Proc. Nat. Acad. Sci.*, **27**, 590 (1941).

<sup>8</sup> van Niel, C. B., Ruben, S., Carson, S. F., Kamen, M. D., and Foster, J. W., *Ibid.*, **28**, 8 (1942).

<sup>9</sup> Rahn, O., *Growth*, **5**, 197 (1941).

<sup>10</sup> Wood, H. G., and Werkman, C. H., *Biochem. Jour.*, **30**, 48 (1936).

<sup>11</sup> Wood, H. G., and Werkman, C. H., *Ibid.*, **32**, 1262 (1938).

<sup>12</sup> Wood, H. G., and Werkman, C. H., *Ibid.*, **34**, 7, 129 (1940).

<sup>13</sup> Wood, H. G., and Werkman, C. H., Hemingway, A., and Nier, A. O., *Jour. Biol. Chem.*, **139**, 483 (1941).

<sup>14</sup> Ruben, S., Kamen, M. D., and Hassid, W. Z., *Jour. Am. Chem. Soc.*, **62**, 3443 (1940).

<sup>15</sup> Allen, M. B., and Ruben, S., *Ibid.* (in press).

## THE ENERGY EQUATION FOR A VISCOUS COMPRESSIBLE FLUID

BY H. J. STEWART

GUGGENHEIM AERONAUTICS LABORATORY, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated April 6, 1942

One of the more important problems discussed in any text book on Dynamic Meteorology is the calculation of the energy of atmospheric motions, and the conversion of potential and thermal energy to the kinetic energy of the motion of the winds. This problem was first discussed by Margules<sup>1</sup> and has since then been discussed by Bjerknes,<sup>2</sup> Brunt<sup>3</sup> and Haurwitz.<sup>4</sup> Rossby has also discussed this work in his lectures on Dynamic Meteorology. It is therefore of considerable interest to note that the development of the energy equation for a viscous compressible fluid by all these latter writers is in error through the omission of terms involving the viscous stresses. On the other hand, this matter is correctly handled in Margules' original paper.

The energy equation for this case can be obtained easily from the dynamical equations of motion for a viscous fluid. This is, using the Cartesian tensor notation,

$$\frac{Du_i}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial x_i} - \frac{\partial V}{\partial x_i} + \frac{1}{\rho} \frac{\partial}{\partial x_j} (\tau_{ij}) \quad (1)$$

where  $x_i$  are rectangular coördinates,  $u_i$  is the fluid velocity vector,  $\rho$  is the fluid density,  $V$  is the gravitational potential and  $\tau_{ij}$  is the viscous stress tensor. Upon taking the scalar product of the velocity vector with this, one obtains, since  $\frac{\partial V}{\partial t} = 0$ ,

$$\rho \frac{D}{Dt} \left\{ \frac{1}{2} u_i u_i + V \right\} = -u_i \frac{\partial p}{\partial x_i} + u_i \frac{\partial}{\partial x_j} (\tau_{ij}). \quad (2)$$

If heat is added to the fluid, the relative energy equation for the element is

$$\frac{Dq}{Dt} = \frac{Di}{Dt} + p \frac{D}{Dt} \left( \frac{1}{\rho} \right) - \frac{1}{\rho} \tau_{ij} \frac{\partial u_i}{\partial x_j} \quad (3)$$

where  $q$  is the heat added per unit mass of fluid, and  $i$  is the internal energy of a unit mass of fluid. The last two terms represent the work done on the external fluid through the deformation of the element. The equation of continuity,

$$0 = \frac{1}{\rho} \frac{D\rho}{Dt} + \frac{\partial u_i}{\partial x_i}, \quad (4)$$

can be used to transform equation (3) to

$$\rho \frac{Dq}{Dt} = \rho \frac{Di}{Dt} + p \frac{\partial u_i}{\partial x_i} - \tau_{ij} \frac{\partial u_i}{\partial x_j}. \quad (5)$$

In order to obtain the complete energy equation, equations (3) and (5) must be added. This gives the energy equation for a viscous compressible fluid as

$$\rho \frac{Dq}{Dt} = \rho \frac{D}{Dt} \left\{ i + V + \frac{1}{2} u_i u_i \right\} + \frac{\partial}{\partial x_i} (p u_i) - \frac{\partial}{\partial x_j} (\tau_{ij} u_i). \quad (6)$$

This equation could have been written directly by the principle of conservation of energy as the last two terms are the work done by the element on the surrounding fluid by pressure and viscous stresses, respectively.

In each of the above mentioned references, the term,  $\tau_{ij} \frac{\partial u_i}{\partial x_j}$ , is omitted from equation (3), and the final energy equation is written as

$$\rho \frac{Dq}{Dt} = \rho \frac{D}{Dt} \left\{ i + V + \frac{1}{2} u_i u_i \right\} + \frac{\partial}{\partial x_i} (p u_i) - u_i \frac{\partial}{\partial x_j} (\tau_{ij}). \quad (7)$$

That this equation is incorrect is easily seen by considering its application to a closed system with fixed walls and to which no heat is added. An auxiliary theorem will first be developed for use in this demonstration. By use of the continuity equation, it can easily be seen that

$$\rho \frac{DF}{Dt} = \frac{\partial}{\partial t} (\rho F) + \frac{\partial}{\partial x_i} (\rho F u_i) \quad (8)$$

where  $F$  is any given quantity.

Thus,

$$\int_V \int \int \rho \frac{DF}{Dt} dv = \frac{\partial}{\partial t} \int_V \int \int \rho F dv \quad (9)$$

with the volume integral extending throughout the given closed system. The integral of the second term of equation (8) vanishes since  $u_i = 0$  on the boundaries. By applying this result to equation (7), it is seen that

$$\frac{\partial}{\partial t} \{I + P + K\} = - \int_V \int \int \frac{\partial}{\partial x_i} (\rho u_i) dv + \int_V \int \int u_i \frac{\partial}{\partial x_j} (\tau_{ij}) dv \quad (10)$$

where  $I$ ,  $P$  and  $K$  are the total internal, potential and kinetic energies within the system. The first integral on the right-hand side vanishes, and the second can be transformed by integration by parts to

$$\frac{\partial}{\partial t} \{I + P + K\} = - \int_V \int \int \tau_{ij} \frac{\partial u_i}{\partial x_j} dv. \quad (11)$$

Now,  $\tau_{ij} \frac{\partial u_i}{\partial x_j}$  is the kinetic energy dissipated per unit time and volume and turned into heat, so this would indicate that the total energy within the system is diminishing at a rate equal to the rate of dissipation of kinetic energy. This result is in contradiction to the principle of the conservation of energy, and equation (7) must therefore be incorrect. Equation (10) is given by Haurwitz<sup>4</sup> (p. 245) although its significance is not recognized.

If the correct energy equation, equation (6), is applied to the given closed system, it is seen that

$$\frac{\partial}{\partial t} \{I + P + K\} = 0. \quad (12)$$

By the principle of the conservation of energy this is obviously the correct result.

Since  $\tau_{ij} \frac{\partial u_i}{\partial x_j}$  is the kinetic energy converted into heat per unit time and volume, equation (3) can be changed into another form,

$$\frac{Dq'}{Dt} = \frac{Dq}{Dt} + \frac{1}{\rho} \tau_{ij} \frac{\partial u_i}{\partial x_j} = \frac{Di}{Dt} + p \frac{D}{Dt} \left( \frac{1}{\rho} \right), \quad (13)$$

where  $\frac{Dq'}{Dt}$  is the total heat added to the element with one part,  $\frac{Dq}{Dt}$ , being



added from outside and the other part,  $\frac{1}{\rho} \tau_{ij} \frac{\partial u_i}{\partial x_j}$ , being an internal heat source arising from the viscous dissipation. This result has been noted by Lamb<sup>5</sup> (p. 646). The error in equation (7) can thus be considered as arising from neglecting the heat produced by viscous dissipation of the kinetic energy.

<sup>1</sup> Margules, M., *Met. Z.*, 23, 481 (1906).

<sup>2</sup> Bjerknes, V., and collaborators, *Physikalische Hydrodynamik*, Chap. 4, Julius Springer, Berlin, 1933.

<sup>3</sup> Brunt, D., *Physical and Dynamical Meteorology*, Chap. XV, The MacMillan Company, New York, 1934.

<sup>4</sup> Haurwitz, B., *Dynamic Meteorology*, Chap. XII, McGraw-Hill Book Co., Inc., New York, 1941.

<sup>5</sup> Lamb, H., *Hydrodynamics*, 8th ed., Cambridge University Press, Cambridge, 1932.

## SENSITIZATION OF MELANOPHORES BY NERVE CUTTING

BY G. H. PARKER

BIOLOGICAL LABORATORIES, HARVARD UNIVERSITY

Communicated April 7, 1942

For some years past evidence has been accumulating to show that effectors and even neurones on denervation became more sensitive to their ordinary chemical activators than they were before their special nervous connections were severed. This question has been discussed recently by Cannon and Rosenblueth (1937) and has been made the subject of a special lecture by Cannon (1939) who has suggested that such sensitization might well represent a law of denervation. The evidence upon which this view is based has been gathered from the responses of denervated smooth muscles, glands, skeletal muscles and isolated neurones to their normal activators. Within a year similar evidence from another category of effectors, the chromatophores, has been presented by Smith (1941) who has shown that the melanophores of a fish, the tautog, can also be sensitized by denervation.

Recent studies on the color changes in catfishes have led me to suspect that the melanophores of this creature may also be open to sensitization by nerve cutting and during the last year I have carried out a set of tests to ascertain if this were true or not. The work was begun at the Marine Biological Laboratory, Woods Hole, and completed at the Harvard Biological Laboratories. To both these institutions I am indebted for accommodations and ample facilities. The cost of materials, reagents and the like used in these investigations was met in large part by a grant from the Permanent Science Fund to the officers of which I wish to express my sincere thanks.

Smith's procedure with the tautog was to determine the times necessary for the concentration of pigment in three states of the color cells of this fish when activated by adrenaline. Melanophores were tested in freshly removed scales with the normal nerve-terminals still attached, in scales with fully degenerated terminals and in scales with regenerated terminals. The result of Smith's investigations was to show that the pigment in cells with degenerated terminals became concentrated about twice as fast as in cells with active terminals.

My own method in testing this question was very different from that used by Smith, and in place of a marine fish with scales like the tautog I used a scaleless, fresh-water fish, the common eastern catfish, *Ameiurus nebulosus*. The chromatophore system in this fish consists exclusively of melanophores, micromelanophores in the epidermis and macromelanophores in the derma. Of these the macromelanophores are much more satisfactory for study and are the color cells that will be referred to in the present investigation. The melanophore pigment in *Ameiurus* is concentrated by the action of nerve-fibres whose effective neurohumor appears to be adrenaline. It is dispersed by two agents, intermedine from the pituitary gland and acetylcholine from an appropriate set of nerve-fibres (Parker, 1940). These facts will be of importance in the further steps of this investigation.

The responses of denervated and innervated melanophores in *Ameiurus* to adrenaline can be conveniently studied in caudal bands on this fish. When a single fin ray is cut in the tail of *Ameiurus* to form a caudal band nerves carrying both dispersing and concentrating fibres are severed. After having been cut the severed fibres remain active for several days, in two weeks they are completely degenerated and in about a month they have regenerated. (Parker, 1934, 1941a; Abramowitz, 1935, 1936.) The best period in which to obtain a band of denervated melanophores is in the third week after the severance of the nerves. All these time relations apply to fishes during the summer when the water in which they live is at approximately 20°C. The adrenaline used in these tests was the Parke, Davis & Co. preparation 1:1000 which was diluted with cold-blooded Ringer's solution to the requisite strength (see table 1). The catfishes tested weighed each about 50 g. and received by subcutaneous injection 0.2 cc. of the given concentrations of adrenaline. At the beginning of the tests the fishes were intermediate in tone. The effects on the innervated tails and on their denervated bands are given in the third and fourth columns of table 1. These effects are best seen in the first five minutes after the injection has been made. Strong concentrations of adrenaline (1:5000 and 1:20,000) blanched the fishes quickly, the tails and bands becoming pale simultaneously. The injections of weak concentrations (1:1,000,000 to 1:50,000,000) had no appreciable effect upon the tails and bands within the first

five minutes. At a concentration of 1:1,000,000 the whole fish including the band and tail commonly blanched markedly in about an hour. At greater dilutions, however, no change whatever in tint was seen.

TABLE 1

COLOR RESPONSES OF THE INNERVATED TAILS (AS WELL AS THE BODIES OF THE FISHES) AND OF THE DENERVATED CAUDAL BANDS OF CATFISHES (*Ameiurus*) DURING THE FIRST FIVE MINUTES AFTER THE INJECTION OF DIFFERENT CONCENTRATIONS OF ADRENALINE

Three fishes of intermediate tint were used in each test and each fish, which weighed about 50 g., received 0.2 cc. of the solution of adrenaline. The amounts of adrenaline in milligrams injected per 100 g. of fish are given in the second column of the table. The concentrations were made by diluting with Ringer's solution the Parke, Davis & Co. preparation of adrenaline chloride 1:1000. (1) Palish after 10 minutes. (2) Palish after 15 minutes. (3) Became intermediate in an hour. (4) Became pale in an hour. (5) No change whatever.

CONCENTRATIONS OF ADRENALINE	ADRENALINE MG. PER 100 G. FISH	TINT OF REGIONS IN FIRST 5 MINUTES INNERVATED TAIL	DENERVATED BAND
1:5,000	0.08	Blanched quickly	Blanched quickly
1:20,000	0.02	Blanched quickly	Blanched quickly
1:50,000	0.008	Intermediate (1)	Decidedly pale
1:100,000	0.004	Intermediate (2)	Somewhat pale (3)
1:1,000,000	0.000,4	Intermediate (4)	Intermediate (4)
1:20,000,000	0.000,02	Intermediate (5)	Intermediate (5)
1:50,000,000	0.000,008	Intermediate (5)	Intermediate (5)

The critical concentrations in this set of records were those of medium strength (1:50,000 and 1:100,000). In the stronger of these the tail remained intermediate in tint while the band became decidedly blanched. The melanophore pigment masses in the tail generally were stellate, but in the band they were irregularly roundish. About ten minutes after the injection had been made the pigment masses of the tail melanophores became generally roundish. With the weaker concentration 1:100,000 the tail during the first five minutes remained intermediate but the bands blanched noticeably. In about an hour the bands were again intermediate. Thus of the seven concentrations of adrenaline recorded in table 1 two (1:50,000 and 1:100,000) were effective in the first five minutes in concentrating the pigment in the denervated melanophores but failed to do so in the innervated ones.

Two other sets of catfishes were tested for adrenaline responses on essentially the same plan as that outlined in table 1. In both these sets the concentrations began at 1:5000; in one they ran to 1:40,000,000 and in the other to 1:100,000,000. In both sets of fishes the extreme concentrations called forth similar responses in the tails and their associated bands. No differential responses were noted at concentrations stronger than 1:50,000, nor weaker than 1:200,000. Five of the six fishes tested with concentration 1:50,000 responded differentially to this solution as did those recorded in table 1. In the sixth fish both the band and tail blanched indistinguish-

ably. To the concentration 1:200,000 (0.002 mg. of adrenaline per 100 g. of fish) the bands in three fishes responded slightly by advanced blanching thus increasing the range of this effect beyond what is shown in table 1. Aside from these minor differences, however, all three sets of tests were in reasonable agreement in that extreme concentrations of adrenaline either strong or weak influenced the melanophores both denervated and innervated in the same way, whereas the intermediate concentrations (1:50,000 to 1:200,000) called forth a differential response in which the denervated color cells were seen to be more sensitive to adrenaline than the innervated ones. Thus these tests on *Ameiurus* support Smith's results (1941) on the sensitization of the melanophores of the tautog by denervation and afford among chromatophores another example of this type of response.

In addition to adrenaline two other important neurohumors, as already mentioned, take part in the activation of catfish melanophores, acetylcholine and intermedine. Both these are darkening agents. Sensitization of smooth muscle to acetylcholine by nerve cutting has already been noted (Cannon and Rosenblueth, 1937) and it is therefore possible that this phenomenon might be met with in *Ameiurus* melanophores. It was this possibility in fact that led me to suggest (Parker, 1941b) what might be an explanation of the sensitization of denervated catfish melanophores to adrenaline. When a caudal band is cut in a catfish both cholinergic and adrenergic fibres are severed and finally degenerate. Consequently in time such a band would become deficient in the two opposing neurohumors, acetylcholine and adrenaline. In the absence of acetylcholine, therefore, the denervated melanophores would be more freely open to the action of adrenaline than when its opponent neurohumor was present. Hence denervation by reducing acetylcholine might render the melanophores more sensitive to adrenaline. Acetylcholine, however, is limited in its action on catfish melanophores (Osborn, 1938; Parker, 1940). It can convert the pale phase into the intermediate one but not the intermediate one into the dark. Thus it acts over only about half the range of the melanophore change.

Notwithstanding this limitation acetylcholine was tested on palish catfishes with bands that had been previously eserized (Parker, 1940). The concentrations of acetylcholine used ran from 1:1000 to 1:100,000,000 and each fish received 0.2 cc. of the solution with which it was tested. The weaker concentrations were without effect and the stronger ones proved poisonous as was evidenced by the death of many of the fishes. Only at 1:100,000 or 0.004 mg. of acetylcholine per 100 g. of fish did the fishes darken and continue to live. At this concentration, which is about one-hundredth of that used by Kiel and Root (1941) in tests on smooth muscle in the cat, no noticeable differential response could be detected between the bands and the rest of the tail. This occurred so regularly that I was led to

conclude that either acetylcholine was a substance to which the melanophores could not be sensitized by denervation or that the technique herein employed was insufficient to bring out this capacity if it was present.

The second darkening neurohumor in the catfish is the pituitary product, intermedine. The injection of aqueous extracts of the pituitary gland of the catfish are well known to darken this animal (Parker, 1934; Abramowitz, 1936). This fish will also darken to injections of Parke, Davis & Co.'s preparation of obstetrical pituitrin which contains intermedine. Groups of three catfishes of intermediate tint and with fully denervated caudal bands were injected each with a given concentration of pituitary extract made from the glands of the catfish itself. These glandular extracts were made by grinding to great fineness in a known volume of Ringer's solution a definite number of pituitary glands taken freshly from the fishes, and then diluting the mixture till the requisite degree was reached. The volume of extract injected in each fish was 0.2 cc. and the concentration of pituitary tissue in each dose injected in terms of glands varied from two to one-twentieth. After the injections were made the bands and tails of the fishes were watched closely for the next few hours. The significant changes began to show within a very short time. When the three fishes, intermediate in tint, were injected with the most concentrated extract, two glands per dose, all began to darken within five minutes. At ten minutes their bands were noticeably darker than the rest of their tails. This condition remained constant for from three to four hours after which both bands and tails were indistinguishably dark and of the same deep tint as the rest of the fish. This condition was maintained for fully a day whereupon the fishes gradually blanched from having been kept in a white, illuminated vessel. When the concentration injected was one gland per fish the changes were essentially the same as those observed with the strongest extract. For the first few hours after injection the bands were clearly darker than their surroundings after which all were of the same deep tone. When the concentration was half a gland per dose the fishes darkened slowly. For about an hour the bands darkened more rapidly than the tails. Later both bands and tails were equally dark. When only a tenth of a gland was injected no certain difference could be distinguished between the bands and the tails, both of which, however, darkened. After injecting one-twentieth of a gland into each fish no change of any kind was noticeable but all three fishes remained of intermediate tint in close agreement with a group of uninjected, intermediate catfishes held as a check. From these records it appears clear that the denervation of catfish melanophores renders them more sensitive to their own intermedine than they were when innervated and that this newly acquired state is comparable to that of adrenaline sensitization after nerve cutting. Experiments of much the kind as those detailed in this paragraph have already been carried out

by other workers (von Frisch, 1911; Abramowitz, 1936), but these investigators were more interested in the final darkening of the fishes than in the steps by which this concluding stage was attained and consequently gave no attention to the changes that have proved significant in the present research.

Other ways of subjecting catfishes with innervated and denervated melanophores to their own intermedine than the one here described were also tried. It is well known that enucleation will induce catfish in the light to become very dark through a discharge of intermedine. Fishes with bands of appropriate age were therefore deprived of their eyes and kept in the light. They quickly darkened in color but the bands darkened indistinguishably from the tails. The failure of the bands to show an advanced darkening was probably due to the rapidity with which under these circumstances the dark phase was attained. I also failed to obtain darker bands than tails by using obstetrical pituitrin as a dispersing agent. After the injection of this extract the fishes darkened but not in a differential way. This extract which is mammalian in source is very probably less reliable as a dispersing agent than the fresh one made directly from the pituitary glands of the catfishes themselves. When hypophysectomized catfishes with seasoned caudal bands were injected with a pituitary extract of one-gland strength they darkened as normal fishes with pituitary glands did, in that their caudal bands darkened more deeply than the rest of the tail and in advance of it. This is what should be expected if the account of the color reactions of the catfish thus far given is correct. From the results here detailed it appears that denervation sensitizes catfish melanophores to intermedine in much the same way that it sensitizes these color cells to adrenaline.

Although catfish melanophores can be sensitized by nerve cutting to intermedine, the fact that they cannot be so sensitized to acetylcholine is strong evidence against the explanation of melanophore sensitization that I have offered elsewhere (Parker, 1941*b*). That explanation depended on the local opposition of one neurohumor to another, acetylcholine to adrenaline. But now that it has been shown that acetylcholine is apparently not involved in sensitization this explanation, even limited as it is to fish melanophores, lacks application. Melanophore sensitization due to nerve cutting like that seen in other sensitizable effectors is so far as I know without obvious explanation.

*Summary.*—1. Denervation of melanophores in the catfish *Ameiurus nebulosus* will sensitize these color cells to appropriate strengths of adrenaline (0.002 to 0.008 mg. of adrenaline per 100 g. of fish). This result supports Smith's conclusion concerning the sensitization of the melanophores in the tautog.



2. Denervated catfish melanophores were not found to be sensitized to acetylcholine.
3. Catfish melanophores were sensitized by nerve cutting to intermediate whose strength of injection was two pituitary glands to half a gland per fish. These were not sensitized to weaker extracts.
4. No general explanation was found for effector sensitization by nerve cutting.

Abramowitz, A. A., "Regeneration of Chromatophoral Nerves," *Proc. Nat. Acad. Sci.*, **21**, 137-141 (1935).

Abramowitz, A. A., "Physiology of the Melanophore System in the Catfish, *Ameiurus*," *Biol. Bull.*, **71**, 259-281 (1936).

Cannon, W. B., "A Law of Denervation," *Amer. Jour. Med. Sci.*, **198**, 737-750 (1939).

Cannon, W. B., and Rosenblueth, A., *Autonomic Neuro-Effector Systems*, New York, 1937, 229 pp.

von Frisch, K., "Beiträge zur Physiologie der Pigmentzellen in der Fischhaut," *Arch. ges. Physiol.*, **138**, 319-387 (1911).

Keil, F. C., and Root, W. S., "Parasympathetic Sensitization in the Cat's Eye," *Amer. Jour. Physiol.*, **132**, 437-445 (1941).

Osborn, C. M., "The Role of the Melanophore-Dispersing Principle of the Pituitary in the Color Changes of the Catfish," *Jour. Exp. Zool.*, **79**, 309-330 (1938).

Parker, G. H., "Color Changes of the Catfish *Ameiurus* in Relation to Neurohumors," *Ibid.*, **69**, 199-233 (1934).

Parker, G. H., "On the Neurohumors of the Color Changes in Catfishes and on Fats and Oils as Protective Agents for Such Substances," *Proc. Amer. Philos. Soc.*, **83**, 379-409 (1940).

Parker, G. H., "Melanophore Bands and Areas Due to Nerve Cutting, in Relation to the Protracted Activity of Nerves," *Jour. Gen. Physiol.*, **24**, 483-504 (1941a).

Parker, G. H., "Hypersensitization of Catfish Melanophores to Adrenaline by Denervation," *Biol. Bull.*, **81**, 302 (1941b).

Smith, D. C., "The Effect of Denervation upon the Response to Adrenalin in the Isolated Fish Scale Melanophores," *Amer. Jour. Physiol.*, **132**, 245-248 (1941).

---

## ON GENERALIZED BIORTHOGONAL EXPANSIONS IN METRIC AND UNITARY SPACES

BY Y. Y. TSENG

DEPARTMENT OF MATHEMATICS, TSING-HUA UNIVERSITY, KUNMING, CHINA

Communicated April 7, 1942

1. *Introduction.*—Recently R. P. Boas, Jr., enunciated in these PROCEEDINGS<sup>1</sup> a principle of biorthogonal expansion, from which follow readily various well-known special expansion theorems. The object of the present note is to give new conditions for expansions, to generalize and strengthen some of the principal results on general expansions occurring in the litera-

ture. This is done by introducing for study certain closely and directly related functional transformations and infinite matrices, with the aid of E. H. Moore's second theory of general analysis (*G. A.*)<sup>2</sup> and a few facts newly encountered there.

Unless specified otherwise, we consider throughout this note (1) *an arbitrarily fixed (semi-definite) positive Hermitian matrix*<sup>2</sup>  $E_*$  on  $PP$ , with "coördinates"  $E_*(p'p'')$  where  $p'p''$  run independently over the *unrestricted* non-empty set  $P$ , and (2) *a general system of functions*  $\{g_p\}$  in a metric space  $\mathfrak{R}$  or a unitary space  $\mathfrak{M}$  (both being complete and left-linear over the real, complex or *quaternionic* numbers), where  $\{g_p\}$  is *not* assumed linearly independent. It forms a biorthogonal system in our generalized sense, briefly a  $B_*$ -system,<sup>3</sup> if, relative to  $E_*$  there exists a system of bounded linear functionals  $\{\psi_p\}$  such that

$$\psi_{p'}(g_{p''}) = E_*(p''p') \quad (p'p'' \text{ in } P).$$

In the sequel it is fundamental that in any complete linear metric space the Cauchy condition on the existence of limit holds for any directed system of elements.<sup>4</sup>

A detailed account of these results and a discussion of milder conditions as well as of other expansions defining oblique projectors will be published elsewhere.

2. *Expansions in Metric Spaces.*—Let  $\mathfrak{R}$  be a complete linear metric space, and  $\{f_p\}$ ,  $\{g_p\}$  be arbitrary systems of elements in  $\mathfrak{R}$ . Then  $\{f_p\}$  is said to be completely accordant to  $\{g_p\}$  or completely  $g$ -accordant, if, for any bounded linear functional  $L$  on  $\mathfrak{R}$ ,

$$L(g_p) = 0 \ (p) \quad \text{implies} \quad L(f_p) = 0 \ (p).$$

We call  $\{f_p\}$  bounded relative to  $\{g_p\}$  or  $g$ -bounded, in case there exists a positive constant  $M$  such that

$$\left\| \sum_{p \in \sigma} a_p f_p \right\| \leq M \cdot \left\| \sum_{p \in \sigma} a_p g_p \right\|$$

for every finite subset  $\sigma$  of  $P$  and arbitrary constants  $\{a_p\}$ .

Relative to  $E_*$  we define a functional integral  $J_*$  provided the following limit exists:

$$J_* a_p f_p \equiv \lim_{\sigma} \sum_{p' \in \sigma} \sum_{p'' \in \sigma} a_{p'} R_*^{\sigma}(p'p'') f_{p''},$$

where  $R_*^{\sigma}$  is Moore's general reciprocal<sup>2</sup> of the finite square matrix  $E_*(p'p'')$  as  $p'p''$  vary independently over  $\sigma$ , and the limit as  $\sigma$  swells is the one introduced years ago by Moore in these PROCEEDINGS.<sup>4</sup> Here and below we always write  $\lim$  for strong limit.

By an  $E_*$ -base of  $\mathfrak{R}$  we mean a  $B_*$ -system  $\{g_p\}$  with an adjoint system  $\{\psi_p\}$  in terms of which every  $h$  of  $\mathfrak{R}$  has and equals the strongly con-



vergent expansion  $J_*\psi_p(h)g_p$ . Observe that relative to an  $E_*$ -base, the expansions are always unique with  $E^*$ -accordant<sup>2</sup> coefficients.

With these preliminaries we can now state:

**THEOREM A.** *Suppose that  $\{g_p\}$  and  $\{f_p\}$  are mutually bounded. Then  $\{g_p\}$  is a completely  $f$ -accordant  $E_*$ -base of  $\mathfrak{R}$  if and only if  $\{f_p\}$  is a completely  $g$ -accordant  $E_*$ -base of  $\mathfrak{R}$ .*

**COROLLARY ( $\alpha$ ).** (Generalization of Boas's theorem.) *If for any set of arbitrary constants  $\{a_p\}$*

$$\left\| \sum_{p \in \sigma} a_p (f_p - g_p) \right\| \leq \lambda \cdot \left\| \sum_{p \in \sigma} a_p f_p \right\|, \quad 0 < \lambda < 1,$$

*then  $\{g_p\}$  is an  $E_*$ -base of  $\mathfrak{R}$  whenever  $\{f_p\}$  is such.*

**COROLLARY ( $\beta$ ).** (First generalization of Paley-Wiener's theorem.<sup>5</sup>) *Let  $\{f_p\}$ ,  $\{g_p\}$  be systems in a complete unitary space, satisfying the condition of Corollary ( $\alpha$ ). Then  $\{g_p\}$  is an  $E_*$ -base, when and only when  $\{f_p\}$  is such.*

Here, the usual proof by iteration<sup>1, 5</sup> breaks down, because such a positive constant  $\lambda$  less than one is no longer available. We note, however, that if  $\{f_p\}$  is an  $E_*$ -base with an adjoint system  $\{\phi_p\}$ , the integral  $J_*\phi_p(h)g_p$  converges strongly and defines a bounded linear transformation  $Ah$  on  $\mathfrak{R}$ . Moreover, it has a unique bounded classical reciprocal  $A^{-1}$  on  $\mathfrak{R}$ . Consequently the adjoint  $(A^{-1})^*$  is also bounded. Then, defining the functionals  $\{\psi_p\}$  to be  $(A^{-1})^*\phi_p$ , we see that  $\{g_p\}$  is a  $B_*$ -system with  $\{\psi_p\}$  as an adjoint system, and obtain for arbitrary  $h$  the strongly convergent expansion  $J_*\psi_p(h)g_p$ .

**3. Expansions in Unitary Space.**—When an inner product is present, sharper results and explicit formulae can be established. Hereafter,  $\mathfrak{M}_g$  denotes the closure of  $\{g_p\}$  (i.e., the closed linear subspace determined by these elements);  $P_g$ , the orthogonal projection on  $\mathfrak{M}_g$ ;  $E_g$ , the Gramian matrix  $(g_{p'}, g_{p''})$  of  $\{g_p\}$ ; and  $J_g a_p h_p$ , the strong limit analogous to  $J_* a_p g_p$  with  $E_*$ ,  $\{g_p\}$  replaced by  $E_g$ ,  $\{h_p\}$ , respectively. We call the unique system  $\{h_p\}$  in  $\mathfrak{M}_g$  such that  $(g_{p'}, h_{p''})$  equals  $E_*(p'p'')$ , the  $E_*$ -adjoint system to  $\{g_p\}$ .

With the help of new matricial results in *G. A.*, recent results on systems of functions,<sup>6</sup> and several lemmas some of which generalize theorems of A. J. Pell<sup>7</sup> and S. Lewin,<sup>8</sup> we derive the consequences collected in the next theorem, where, it is to be noted, the essential restriction of "nearness"<sup>1, 5</sup> as measured by  $0 < \lambda < 1$  is entirely discarded.

**THEOREM B.** *In order that the  $g_p$ 's form a  $B_*$ -system, the integral  $J_* c_p g_p$  converge weakly for every  $E_*$ -modular<sup>2</sup>  $\{c_p\}$ , and every  $g$  of  $\mathfrak{M}_g$  have such a weak expansion, it is necessary and sufficient that  $E_*$  and  $E_g$  be mutually modular.<sup>2</sup>*

*If so, the expansion is unique and converges strongly. The closures of*

$\{g_p\}$  and its unique  $E_*$ -adjoint system  $\{h_p\}$  coincide;  $E_*$  and  $E_h$  are mutually modular; and

$$c_p = J_g(g, g_{p'})E_*(p'p) = J_*(g, g_{p'})E_g^{-1}(p'p),$$

$$J_*(f, h_p)g_p = P_g f = J_*(f, g_p)h_p.$$

Moreover, we have the sharpened and generalized Paley-Wiener inequalities:<sup>9,2</sup>

$$\frac{1}{\overline{M}_*(E_g)} J_*(f, g_p)(g_p, f) \leq \|P_g f\|^2 \leq \frac{1}{*M(E_g)} J_*(f, g_p)(g_p, f),$$

$$\underline{M}_*(E_g) \cdot J_*(f, h_p)(h_p, f) \leq \|P_g f\|^2 \leq \overline{M}_*(E_g) \cdot J_*(f, h_p)(h_p, f).$$

Lack of space prevents describing further expansions closely attached to the above situation. From this theorem we deduce the following result which has much contact with the recent literature.

**THEOREM C.** *Given two non-zero positive constants  $d, e$  with  $d$  less than  $e$ . For the existence of a system  $\{f_p\}$  in  $\mathfrak{M}$  such that*

$$(1a) \quad (f_{p'}, f_{p''}) = e^2 \cdot E_*(p'p''),$$

$$(1b) \quad \left\| \sum_{p \in \sigma} a_p (g_p - f_p) \right\|^2 \leq d^2 \cdot \sum_{p' \in \sigma} \sum_{p'' \in \sigma} a_{p'} E_*(p'p'') \overline{a_{p''}}$$

*it is necessary and sufficient that one of the equivalent sets of conditions (2) and (3) be satisfied:*

$$(2a) \quad \text{the rows of } E_* \text{ are modular as to } E_g,$$

$$(2b) \quad \frac{1}{(e+d)^2} J_*(f, g_p)(g_p, f) \leq \|P_g f\|^2 \leq \frac{1}{(e-d)^2} J_*(f, g_p)(g_p, f);$$

$$(3a) \quad E_* \text{ and } E_g \text{ are mutually modular,}$$

$$(3b) \quad (e-d)^2 \leq \underline{M}_*(E_g) \leq \overline{M}_*(E_g) \leq (e+d)^2.$$

In fact, any system  $\{f_p\}$  fulfilling conditions (1a) (1b) and such that one of the spaces  $\mathfrak{M}_f, \mathfrak{M}_g$  lies in the other must have  $\mathfrak{M}_f$  coincident with  $\mathfrak{M}_g$ . Again, the closures of  $\{g_p\}$  and its unique  $E_*$ -adjoint  $\{h_p\}$  are equal. We have

$$\underline{M}_*^2(E_g) \cdot J_g(f, h_p)(h_p, f) \leq \|P_g f\|^2 \leq \overline{M}_*^2(E_g) \cdot J_g(f, h_p)(h_p, f).$$

Finally, the strongly convergent expansions, Paley-Wiener inequalities and other results in Theorem B all remain valid.

**COROLLARY ( $\alpha$ ).** (Second generalization of Paley-Wiener's theorem.) In particular, the theorem holds when  $d = \lambda < 1, e = 1$ .

**COROLLARY ( $\beta$ ).** (Generalization of Duffin-Eachus' theorem.<sup>10</sup>) Let  $E_*$  be a row-finite matrix, and  $\{g_p\}$  a "minimal system" in which the omission of any term renders the closure of the remaining system a proper subspace of  $\mathfrak{M}_g$ . Then condition (2b) is both necessary and sufficient for (1a) (1b).

Concerning this theorem and its corollaries, a few remarks may serve to clarify the situation.

*Remarks.*—(α) Counter-examples can be constructed to show that, unlike the ones stated in this and the preceding theorems, the inequalities as originally given by Paley and Wiener in their important closure theorem are not the best possible. (β) In the cited abstract Duffin and Eachus did not mention any additional requirement beside (2b). But, in any pair of systems  $\{f_p\}, \{g_p\}$  satisfying conditions (1a) (1b) with  $E_*$  equal to the (infinite) identity matrix  $I$ , the system  $\{g_p\}$  must always be minimal. And, there is a linearly independent system  $\{g_p\}$  having property (2b), which, nevertheless, fails to be minimal even when  $E_* = I$ . (γ) The restriction  $d < e$  is really not so narrow as it may seem. Indeed, the theorem applies to the last example of Lewin<sup>8</sup> and supplies more information.

As to the proof, it is in the part (2)  $\rightarrow$  (1) that we find a point of special interest. Let  $\{w_p\}$  be a complete system for  $\mathfrak{M}_k$  with its Gramian equal to  $E_*$  and  $K_k$  the matrix whose "coördinates"  $K_k(p_1 p_2)$  are  $(g_{p_1}, w_{p_2})$ . Then,  $K_k$  possesses a classical right reciprocal, since  $J_* K_k^* K_k, J_* K_k K_k^*$  each has lower  $E_*$ -modulus<sup>2</sup> greater than zero by virtue of a previous result on "Schmidt's problem."<sup>6</sup> As in the theory of linear transformations,<sup>11</sup> this fact in turn permits us to write  $K_k$  in its right polar form

$$K_k = J_* S_k U_k, S_k = \sqrt{E_k}$$

where  $S_k$  is (definite) positive Hermitian and  $U_k$  unitary. If we now set

$$f_p \equiv e \cdot J_* U_k(p p_1) w_{p_1},$$

we can show that this system has all the required properties in (1) by reference to some matricial results of G. A.

<sup>1</sup> These PROCEEDINGS, 26, 139–143 (1940).

<sup>2</sup> This work of Moore, entitled *General Analysis*, is being published in *Memoirs of the American Philosophical Society*, 1935– . The less familiar notions of G. A. here utilized are the following: (a) A Hermitian matrix  $E_*$  is (semi-definite) *positive*, in case every principal minor determinant of finite order has non-negative value. (b) Let the canonical form of a Hermitian matrix  $H$  be  $H = \sum_m r_m I_m$  where  $r_m$  is a real characteristic value of  $H$ , and  $I_m$  the idempotent Hermitian matrix associated with  $r_m$ . Then we call the following matrix the *general reciprocal* of  $H$ :  $R_m \equiv \sum_m 1/r_m' I_m$ , the prime indicating that the fraction is taken to be zero when the denominator vanishes. (c) We call a system of constants  $\{c_p\}$  *accordant* to  $E_*$  in case

$$\sum_{p' \in \sigma} \sum_{p'' \in \sigma} a_p E_*(p' p'') \bar{a}_{p''} = 0 \text{ implies } \sum_{p \in \sigma} c_p \bar{a}_p = 0.$$

(d)  $\{c_p\}$  is  $E_*$ -*modular* in case there exists a constant  $N$  such that

$$|\sum_{p \in \sigma} c_p \bar{a}_p|^2 \leq N \cdot \sum_{p' \in \sigma} \sum_{p'' \in \sigma} a_p E_*(p' p'') \bar{a}_{p''}$$

for every system of constants  $\{a_p\}$ . (e) A Hermitian matrix  $H$  is  $E_*$ -modular, if we can find a constant  $M$  with

$$\left| \sum_{p' \in \sigma} \sum_{p'' \in \sigma} \bar{a}_{p'} H(p' p'') a_{p''} \right| \leq M \cdot \sum_{p' \in \sigma} \sum_{p'' \in \sigma} a_{p'} E_*(p' p'') \bar{a}_{p''}.$$

The lower and upper *moduli*, denoted by  $\underline{M}_*(H)$ ,  $\bar{M}_*(H)$  are, respectively, the lower and upper bounds of  $J_* J_* c_p H(p' p'') c_{p''}$  for all  $E_*$ -modular systems  $\{c_p\}$  with  $J_* c_p \bar{c}_p \leq 1$ . Cf. Riesz, F., *Szeged Acta*, 5, 23-54 (1930).

<sup>1</sup> Cf. Radon, J., *Wiener Berichte*, 122 (IIa), 1295-1438 (1913); Banach, S., *Théorie des Opérations Linéaires*, 1932; Kaczmarz, S., and Steinhaus, H., *Theorie der Orthogonalreihen*, 1935. The convergence theorems of Banach can be extended to denumerable  $B_*$ -systems.

<sup>2</sup> In other words, a complete linear metric space is also "hyper-complete." By a directed system we mean one in which there is a transitive relation with composition property. See Moore, these PROCEEDINGS, 1, 628-632 (1915); Birkhoff, G., *Ann. Math.*, 38, 39-56 (1937).

<sup>3</sup> *Fourier Transforms in the Complex Domain*, 1934, pp. 100-106.

<sup>4</sup> We make frequent use of results of two papers by Tseng in the *Science Reports of Tsing-Hua Univ.*, Series A: "On Schmidt's problem . . .," 3, 299-316 (1936); "Expansions according to an arbitrary system . . .," forthcoming in Vol. 4, No. 4.

<sup>5</sup> Pell, *Trans. Amer. Math. Soc.*, 12, 135-164 (1911).

<sup>6</sup> Lewin, *Math. Ztschr.*, 32, 491-511 (1930).

<sup>7</sup> Cf. Remark ( $\alpha$ ) to Theorem C.

<sup>8</sup> Duffin, R. J., and Eachus, J. J., *Bull. Amer. Math. Soc.*, abstract no. 265, 46, 415 (1940).

<sup>9</sup> Stone, M. H., *Linear Transformations in Hilbert Space*, 1932, p. 332.

## REPRESENTATION OF ONE-PARAMETER SEMI-GROUPS OF LINEAR TRANSFORMATIONS

BY EINAR HILLE

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY

Communicated April 7, 1942

1. Let  $E$  be a separable normed complete linear vector space. Let  $T_s$  be defined for  $s > 0$  as a linear bounded transformation on  $E$  to  $E$  such that

$$T_s T_t = T_t T_s = T_{s+t}, \quad s > 0, t > 0. \quad (1.1)$$

We suppose further:

- (1)  $T_s$  is weakly measurable for  $s > 0$ ,
- (2)  $\|T_s\| \leq 1, s > 0$ ,
- (3)  $T_s(E)$  is dense in  $E$ .

Here (2) can be replaced by  $\|T_s\| \leq M$  for  $0 < s < 1$  without inconvenience but with some unessential modifications of the results. It is enough that (3) holds for a single  $s$ , it will then hold for all. If (3) is not satisfied in the

original space, we can often find a subspace where it holds and the discussion then applies to that subspace.<sup>1</sup>

On the basis of these assumptions we have proved strong continuity,<sup>2</sup> differentiability and representation theorems for semi-groups. The prototype of such results is the theorem of M. H. Stone on unitary groups in Hilbert space.<sup>3</sup> Our method is essentially that of Stone. Corresponding results for groups on spaces of type (B) have been announced by I. Gelfand and M. Fukamiya.<sup>4</sup>

We form

$$R(\lambda)x = -\int_0^\infty e^{-\lambda s} T_s x ds, \quad x \in E, \Re(\lambda) > 0 \quad (1.2)$$

where the integral exists for instance in the sense of Bochner and Dunford.<sup>5</sup> This formula defines a linear transformation  $R(\lambda)$  on  $E$  to  $E$  for  $\Re(\lambda) > 0$  and  $\|R(\lambda)\| \leq [\Re(\lambda)]^{-1}$ . The range  $R(\lambda)(E)$  of  $R(\lambda)$  is dense in  $E$  for every  $\lambda$ . Further  $(\lambda - \mu)R(\lambda)R(\mu) = R(\lambda) - R(\mu)$  for  $\Re(\lambda) > 0$ ,  $\Re(\mu) > 0$  and  $R(\lambda_0)x = 0$  implies  $x = 0$ .

$R(\lambda)$  is the resolvent of a closed linear transformation  $A$  whose domain of definition,  $D(A)$ , say, includes  $R(\lambda)(E)$  for every  $\lambda$  and is, therefore, dense in  $E$ . We have

$$R(\lambda)(A - \lambda I) = I \text{ in } D(A), (A - \lambda I)R(\lambda) = I \text{ in } E, \quad (1.3)$$

and for every  $x \in D(A)$  we have in the sense of strong convergence

$$Ax = \lim_{h \rightarrow 0} A_h x, \quad A_h = \frac{1}{h} [T_h - I]. \quad (1.4)$$

The point spectrum of  $A$  may be dense in  $\Re(\lambda) \leq 0$ . From the fact that  $D(A)$  is dense in  $E$  we conclude that for  $h \rightarrow 0$

$$T_h x \rightarrow x, \quad x \in E, \quad (1.5)$$

again in the sense of strong convergence.

2. With the aid of (1.2) we can derive a number of representations of the semi-group. The classical inversion formula of Laplace-Stieltjes integrals can be made to yield the result that for  $c > 0$ ,  $s \geq 0$  and  $x \in E$

$$\int_0^s T_u x du = -\lim_{\omega \rightarrow \infty} (2\pi i)^{-1} \int_{c-i\omega}^{c+i\omega} e^{s\lambda} R(\lambda) x \frac{d\lambda}{\lambda}, \quad (2.1)$$

where the limit exists in the strong sense. If  $x \in D(A)$ , the integral obtained by letting  $\omega \rightarrow \infty$  is convergent. Further, for  $s > 0$ ,  $x \in D(A)$

$$T_s x = -\lim_{\omega \rightarrow \infty} (2\pi i)^{-1} \int_{c-i\omega}^{c+i\omega} e^{s\lambda} R(\lambda) x d\lambda \quad (2.2)$$

in the strong sense. For  $s = 0$ , the formula gives  $1/2x$  instead of  $x$ . If

the limit in (2.2) be replaced by the  $(C, 1)$ -limit, we can omit the restriction  $x \in D(A)$ .

The right-hand side of formula (2.2) can be interpreted as a definition of  $\exp (sA)x$  in the sense of operational calculus,<sup>6</sup> so that

$$T_s x = \exp (sA)x, \quad x \in D(A), \quad s > 0. \quad (2.3)$$

Other interpretations of  $\exp (sA)$  are obtainable by our methods. Thus

$$\int_0^s T_u x du = \lim_{h \rightarrow 0} \int_0^s \exp (uA_h)x du \quad (2.4)$$

for  $s \geq 0$  and every  $x \in E$ , while

$$T_s x = \lim_{h \rightarrow 0} \exp (sA_h)x \quad (2.5)$$

for  $x \in D(A)$ . Here

$$\exp (sA_h)x = \sum_0^\infty \frac{s^n}{n!} A_h^n x. \quad (2.6)$$

The proof of these relations is obtained by observing that  $\exp (sA_h)$  defines a semi-group for  $s > 0$ , the corresponding differential operator being  $A_h$  with a resolvent  $R_h(\lambda)$ , and the latter can be shown to converge uniformly to  $R(\lambda)$ , sufficiently regularly with respect to  $\lambda$ , so that if  $R(\lambda)$  is replaced by  $R_h(\lambda)$  in (2.1), we can pass to the limit with  $h$  under the sign of integration.<sup>7</sup>

For the case of a group I. Gelfand has proposed the interpretation

$$\exp (sA)x = \sum_0^\infty \frac{s^n}{n!} A^n x. \quad (2.7)$$

Gelfand's method appears to break down for semi-groups and it is not clear, in general, that there will exist elements belonging to all sets  $D(A^n)$  for which the series is convergent in any sense whatsoever. For special semi-groups the situation may be different. Thus if  $T_s$  is defined on a Lebesgue space and  $T_s$  commutes with real translations on the point variable, then it is an easy matter to show that there exists a subspace  $E_0$  dense on  $E$  such that the series in (2.7) converges for every  $x \in E_0$  and every finite real or complex  $s$  in the sense that the sum of the norms of the terms is convergent.

<sup>1</sup> Condition (3) is used in proving that  $R(\lambda)(E)$  and hence also  $D(A)$  are dense in  $E$ . We also use the separability of the space at this point.

<sup>2</sup> Strong continuity for  $s > 0$  but not for  $s = 0$  has been proved by Dunford, N., "On One-Parameter Groups of Linear Transformations," *Ann. Math.*, ser. 2, 39, 569-573 (1938), under more general assumptions.

<sup>3</sup> These PROCEEDINGS, 16, 173-174 (1930), and *Ann. Math.*, ser. 2, 33, 643-648 (1932).

<sup>4</sup> Gelfand, I., "On One-Parametrical Groups of Operators in a Normed Space," *Compt. Rend. Acad. Sci. U. R. S. S.*, 25, 713-718 (1939); Fukamiya, M., "On One-Parameter Groups of Operators," *Proc. Imp. Acad. Tokyo*, 16, 262-265 (1940). Fukamiya assumes separability; his proof, partly somewhat fragmentary, uses the method of

Stone, but does not go beyond the existence of  $A$  and  $R(\lambda)$  and stops short of representation theorems. Gelfand in part uses an unpublished notion of integration of his own; he claims to be able to dispense with separability, and arrives at the representation (2.7) below, valid in a subspace dense in  $E$ .

<sup>5</sup> Bochner, S., "Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind," *Fund. Math.*, **20**, 262-276 (1933); Dunford, N., "Integration in General Analysis," *Trans. Amer. Math. Soc.*, **37**, 441-453 (1935). Separability plus weak measurability implies that the integrand is measurable in the sense of Bochner (see Pettis, B. J., "Integration in Vector Spaces," *Trans. Amer. Math. Soc.*, **44**, 277-304 (1938), (Theorem 1.1 and Corollary 1.11) and the norm is measurable and dominated by an integrable function.

<sup>6</sup> Cf. Stone, M. H., *Linear Transformations in Hilbert Space*, New York, 1932, Chapter VI.

<sup>7</sup> The author has found a simpler proof of (2.5) which is valid for all  $x$  and  $E$ —Condition (3) can be replaced by the weaker condition (3\*): The least linear hull of the range spaces is dense in  $E$ . This condition is necessary for the validity of (1.5). [Added in proof.]

## ON A CLASS OF SURFACES

BY G. FUBINI

THE INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.

Communicated March 30, 1942

An interesting problem of the differential projective geometry is the problem of finding all the surfaces  $S$  which possess at least two systems of curves such that the curves of the same system are projective to each other. Many particular surfaces  $S$  are well known (see below). We can suppose that the two systems of curves on  $S$  are given by the equations  $u = \text{const.}$  and  $v = \text{const.}$  Let us suppose that the line  $\Gamma_v = 0$  defined by  $v = 0$  is given by the equation

$$x_i = b_i(u) \quad (i = 1, 2, 3, 4) \quad (1)$$

in which the  $b$  are functions of  $u$ , and the  $x$  are projective homogeneous coordinates, and that the line  $\Gamma_u = 0$  is given by

$$x_i = e_i(v) \quad (e_i \text{ functions of } v). \quad (2)$$

The point  $A(u = v = 0)$  will be the point

$$x_i = \lambda_i = b_i(0) = e_i(0). \quad (3)$$

If the lines  $v = \text{const.}$  are projective to each other, every line  $v = \text{const.}$  can be deduced from the line  $\Gamma_v = 0$  by means of a projectivity  $T_v$ , the coefficients  $a_{ik}$  of which ( $i, k = 1, 2, 3, 4$ ) are functions of  $v$  such that the determinant of the  $a_{ik}$  is different from zero and that, for  $v = 0$ , the pro-

jectivity  $T_v$  is the identity. Therefore the coördinates  $x$  of the most general point of the surface  $S$  will be given by

$$x_i = \sum_k a_{ik}(v)b_k(U) \quad (4)$$

where

$$U = U(u, v) \quad (5)$$

is a function of  $u, v$  such that for  $v = 0$ ,  $U = u$ . The point  $u = u_0$  of the line  $\Gamma_v = 0$  is carried by the projectivity  $T_v$  into the point

$$x_i = \sum_k a_{ik}(v)b_k(u_0). \quad (6)$$

By comparing (4) and (6) we find that

$$U(u, v) = u_0 = \text{const.} \quad (7)$$

is the equation of the *trajectories* of the projectivities  $T_v$ .

In the same manner, by considering the lines  $u = \text{const.}$ , we shall find another system of projectivities  $T_u$ , the coefficients  $c_{ik}$  of which are functions of  $u$ ; their determinant is different from zero; the projectivity  $T_u$ , for  $u = 0$ , is the identity. Therefore every point of  $S$  can be defined also by the equation

$$x_i = \sum_k c_{ik}(u)e_k(V) \quad (V = V(u, v)) \quad (V = v \text{ for } u = 0) \quad (4)_{\text{bis}}$$

in which  $V$  is a function of  $u, v$ , and the equation  $V = \text{const.}$  determines the trajectories of the projectivities  $T_u$ . Since (4) and  $(4)_{\text{bis}}$  determine the same point, and the  $x$  are homogeneous coördinates, we obtain the following system of equations

$$\sum_k a_{ik}(v)b_k(U) = R(u, v) \sum_h c_{ih}(u)e_h(V) \quad (i = 1, 2, 3, 4) \quad (\text{I})$$

in which  $R$  is a function of  $u, v$  finite and different from zero. It seems very difficult to find the most general solution of (I); and the problem is also difficult even if we suppose that the surface  $S$  possesses a third system of curves projective to each other. Particular solutions are well known: the quadrics, the surfaces of rotation, the surfaces of translation, the ruled surfaces, the asymptotic lines of which are curves belonging to a linear complex, etc. We know that it often happens that the trajectories  $U = \text{const.}$  and  $V = \text{const.}$  are identical with the lines  $u = \text{const.}$ ,  $v = \text{const.}$ , so that equation (I) becomes (by supposing  $U = u$ ,  $V = v$ )

$$\sum_k a_{ik}(v)b_k(u) = R(u, v) \sum_h c_{ih}(u)e_h(v). \quad (\text{II})$$



This note is devoted to the solution of (II) under the hypothesis that neither the line  $v = 0$  nor the line  $u = 0$  belongs to an algebraic surface. Therefore neither the lines  $v = \text{const.}$  nor the lines  $u = \text{const.}$  will belong to an algebraic surface. (It follows that they cannot be straight lines or plane curves.) For this exceptional case other methods are necessary.

According to our hypotheses we can find two constants,  $\alpha \neq 0$ ,  $\gamma \neq 0$ , such that

$$a_{ik}(0) = \alpha \delta_{ik} \quad c_{ik}(0) = \gamma \delta_{ik} \quad (\delta_{ii} = 1; \delta_{ik} = 0 \text{ if } i \neq k).$$

Equation (II) is equivalent to the equation

$$\sum_k \bar{a}_{ik}(v) \bar{b}_k(u) = \bar{R}(u, v) \sum_k \bar{c}_{ih}(u) \bar{e}_h(v)$$

in which

$$\begin{aligned} \bar{a}_{ik} &= \frac{a_{ik}}{\alpha}, & \bar{b}_k(u) &= \alpha \frac{b_k(u)}{R(u, 0)} & \bar{R}(u, v) &= \frac{R(u, v)R(0, 0)}{R(u, 0)R(0, v)} \\ \bar{c}_{ih} &= \frac{c_{ih}}{\gamma} & \bar{e}_h(v) &= \gamma \frac{R(0, v)}{R(0, 0)} e_h(v). \end{aligned}$$

Since the coördinates are homogeneous, the point  $x_k = \bar{b}_k(u)$  is identical with the point  $x_k = b_k$  and the point  $x_k = \bar{e}_k$  is identical with the point  $x_k = e_k$ . We find that

$$\bar{a}_{ik}(0) = \bar{c}_{ih}(0) = \delta_{ik} \quad \bar{R}(0, v) = \bar{R}(u, 0) = \bar{R}(0, 0) = 1.$$

By changing notation, and disregarding the dashes, we obtain the equations

$$\begin{aligned} \sum_k a_{ik}(v) b_k(u) &= R(u, v) \sum_h c_{ih}(u) e_h(v) & (i = 1, 2, 3, 4) \\ \frac{\sum_k a_{ik}(v) b_k(u)}{a_{ik}(0) = c_{ik}(0) = \delta_{ik}} &= \frac{R(u, v) \sum_h c_{ih}(u) e_h(v)}{R(0, v) = R(u, 0) = R(0, 0) = 1} & (8) \end{aligned}$$

and by supposing  $u = v = 0$ , we obtain again  $b_i(0) = e_i(0)$ , so that we can always write equation (3) with new values of the  $\lambda$ . By supposing only  $u = 0$ , or  $v = 0$ , we get

$$\begin{aligned} b_i(u) &= \sum_h \lambda_h c_{ih}(u), \\ e_i(v) &= \sum_k \lambda_k a_{ik}(v). \end{aligned} \tag{9}$$

By means of a projective change of coördinates we could suppose that the point  $u = v = 0$  (the coördinates of which are the  $\lambda_i$ ) is the point  $(0, 0, 0, 1)$ . In this case we might write

$$b_i(u) = c_{i4}(u) \quad e_i(v) = a_{i4}(v).$$

But this simplification is of no importance for what follows. Let us choose

four values of  $u$  ( $u = u_1, u_2, u_3, u_4$ ) and let us consider the determinant  $\Delta_1$  the  $r$ th row of which ( $r = 1, 2, 3, 4$ ) is

$$e_r(v) \quad \sum_h c_{rh}(u_2)e_h(v) \quad \sum_h c_{rh}(u_3)e_h(v) \quad \sum_h c_{rh}(u_4)e_h(v).$$

From (8) and (9) we deduce that it is equal to the determinant, the  $r$ th row of which is

$$\sum_k \lambda_k a_{rk}(v) \quad \frac{1}{R(u_2, v)} \sum_k a_{rk}(v) b_k(u_2) \dots \frac{1}{R(u_4, v)} \sum_k a_{rk}(v) b_k(u_4).$$

Therefore the determinant  $\Delta_1$  is equal to

$$\Delta_1 = \frac{AB_1}{R(u_2, v)R(u_3, v)R(u_4, v)} \quad (10)$$

in which  $A$  is the determinant of the  $\alpha_{ik}$  and

$$B_1 = \begin{vmatrix} \lambda_1 & b_1(u_2) & b_1(u_3) & b_1(u_4) \\ \lambda_2 & b_2(u_2) & b_2(u_3) & b_2(u_4) \\ \lambda_3 & b_3(u_2) & b_3(u_3) & b_3(u_4) \\ \lambda_4 & b_4(u_2) & b_4(u_3) & b_4(u_4) \end{vmatrix}. \quad (11)$$

We recall that  $\lambda_i = b_i(0)$ . Since the curve  $\Gamma_v = 0$  is not plane, we can suppose that  $B_1 \neq 0$ . (We already know that  $A$  and  $\frac{1}{R(u, v)}$  are different from zero, too.)

In the same manner one finds that the determinant  $\Delta$ , the  $r$ th row of which is

$$\sum_h c_{rh}(u_1)e_h(v) \quad \sum_h c_{rh}(u_2)e_h(v) \quad \sum_h c_{rh}(u_3)e_h(v) \quad \sum_h c_{rh}(u_4)e_h(v),$$

is equal to

$$\Delta = \frac{AB}{R(u_1, v)R(u_2, v)R(u_3, v)R(u_4, v)} \quad (12)$$

in which  $B$  is the determinant of the  $b_r(u_s)$ ,

$$B = \begin{vmatrix} b_1(u_1) & \dots & b_1(u_4) \\ \dots & \dots & \dots \\ b_4(u_1) & \dots & b_4(u_4) \end{vmatrix}. \quad (13)$$

From these identities we deduce that

$$\frac{\Delta_1}{\Delta} = \frac{B_1}{B} R(u_1, v) \text{ or } R(u_1, v) = \frac{\Delta_1 : B_1}{\Delta : B}$$

Let us now consider  $u_2, u_3, u_4$  as constants, and  $u_1 = u$  as a variable. We realize without difficulty that  $\Delta_1$  is a homogeneous polynomial  $P(e)$  in  $e_i$  of fourth degree, the coefficients of which are functions of *only*  $u_2, u_3, u_4$  and are therefore constants. Also  $B_1$  is constant. On the other hand,  $\Delta:B$  is a homogeneous polynomial  $P_u(e)$  of fourth degree in  $e_i$ , the coefficients of which are functions of  $u$ . And we have seen that both of these polynomials are different from zero. Therefore

$$R(u, v) = \frac{P(e)}{P_u(e)}. \quad (14)$$

We could arrive at this result in another way. Let us write  $b_{kr} = b_k(u_r)$  and indicate by  $\beta_{kr}$  the algebraic complement of  $b_{kr}$  in the determinant  $B$  of the  $b_k(u_r)$ .

By supposing  $u = u_1, u_2, u_3, u_4$  in (8) we obtain four linear equations in the four unknowns  $a_{i1}, a_{i2}, a_{i3}, a_{i4}$ ; by solving these equations in the usual way, we obtain

$$a_{ij} = \sum_r \beta_{jr} R(u_r, v) \sum_h c_{ih}(u_r) e_h(v)$$

and from (9)

$$e_i(v) = \sum_{j,r} \lambda_j \beta_{jr} R(u_r, v) \sum_h c_{ih}(u_r) e_h(v) \quad (i = 1, 2, 3, 4).$$

I now solve these four equations by considering the

$$z_r = \sum_j \lambda_j \beta_{jr} R(u_r, v) \quad (r = 1, 2, 3, 4)$$

as unknowns. I find (by supposing  $r = 1$ ) that the value of

$$\sum_j \lambda_j \beta_{j1} R(u_1, v) = \frac{B_1}{B} R(u_1, v)$$

is equal to  $\Delta_1:\Delta$ .

In the same way, by interchanging the parameters  $u, v$ , and therefore by interchanging  $R$  and  $1/R$ , I shall find that

$$\frac{1}{R(u, v)} = \frac{Q(b)}{Q_v(b)} \quad (15)$$

in which  $Q$  and  $Q_v$  are homogeneous polynomials of fourth degree in  $b_i$ ; the coefficients of  $Q$  are constants, the coefficients of  $Q_v$  are functions of  $v$ . By comparing (14) and (15) we obtain

$$P_u(e) = \frac{P(e)Q(b)}{Q_v(b)} \text{ or } P_u(e)Q_v(b) = P(e)Q(b). \quad (16)$$

This equation is an identity in  $u, v$  if  $e_i$  and  $b_i$  are the coördinates of a point whatsoever of the curves  $\Gamma_u = 0$  and  $\Gamma_v = 0$ . If I write, for instance,  $x_i = e_i(v)$ , I cannot state that this equation is an identity in  $x_1, x_2, x_3, x_4$  (for every value of  $u$ , because  $Q_v(b)$  is a function of  $v$ ). But it is very easy to consider this equation from a deeper point of view. Let  $\epsilon$  be the coefficients of  $P_u(e)$ , which are functions of  $u$ ; and let us write

$$P_u(e) = \epsilon_{1111}e_1^4 + \epsilon_{1112}e_1^3e_2 + \dots + \epsilon_{1234}e_1e_2e_3e_4 + \dots + \epsilon_{4444}e_4^4 = \frac{P(e)Q(b)}{Q_v(b)}. \quad (17)$$

I consider now 35 values of  $v$ , for instance  $v = v_r$  ( $r = 1, 2, \dots, 35$ ). By supposing  $v = v_r$  in (17), I obtain 35 linear equations in the 35 coefficients and, if the curve  $\Gamma_u = 0$  ( $x_i = e_i(v)$ ) does not belong to an algebraic surface of degree  $m \leq 4$  I can suppose that the  $v_r$  are such that the determinant of the coefficients of the  $\epsilon$  in these 35 equations is different from zero. We can therefore solve these equations with respect to the  $\epsilon$ ; and, by remarking that, for  $v = v_r$ , the last member of (17) is equal to  $Q(b)/K_r(b)$  in which  $K_r$  is a new homogeneous polynomial of fourth degree in the  $b$ , with constant coefficients, I deduce that every  $\epsilon$  is equal to an expression

$$\sum h_r \frac{Q(b)}{K_r(b)} \quad (h_r = \text{const.}).$$

(The constants  $h_r$  change if we change the coefficient  $\epsilon$  which we wish to calculate.) From (17) we deduce consequently that

$$P_u(e) = \sum \pi_r(e) \frac{Q(b)}{K_r(b)} \quad (18)$$

in which  $\pi_r(e)$  is a homogeneous polynomial of degree 4 in  $e$ , with constant coefficients. In the same manner we prove that

$$Q_v(b) = \sum T_s(b) \frac{P(e)}{H_s(e)} \quad (19)$$

in which  $T_s$  and  $H_s$  are homogeneous polynomials of fourth degree (in the  $b$  or in the  $e$ ) with constant coefficients. By substituting (18) and (19) in (16), we obtain

$$\sum \frac{\pi_r(e)}{K_r(b)} \sum \frac{T_s(b)}{H_s(b)} = 1. \quad (20)$$

For every value of  $v$  this equation must be an identity in the  $b$ ; if it were not so, equation (2) would be an algebraic equation in the  $b$ , and the curve  $x = b_i(u)$  would belong to an algebraic surface; which is contrary to our

hypothesis. Let us give to  $v$  a constant value, such that all the  $H_s(e)$  are different from zero. The equation

$$\sum \frac{T_s(b)}{H_s(e)} = 0 \quad (21)$$

defines a surface  $\Sigma$ , if the  $b$  are considered as homogeneous coördinates of a point. And, since (20) is an identity in the  $b$  (for every value of  $v$ ), it is obviously necessary that the points of this surface  $\Sigma$  satisfy the equation

$$0 = K_1(b) \quad K_2(b) \dots \quad (22)$$

(The second member is the product of the denominators of  $\frac{\pi_r(e)}{K_r(b)}$ .) But (22) does not depend on the value of  $v$ , whereas (21) depends on this value. This is possible only if

$$\sum \frac{T_s(b)}{H_s(e)} = \psi(v)L(b) \quad (23)$$

in which  $\psi$  is a function only of  $v$  and  $L$  is a homogeneous polynomial of fourth degree in  $b$  with constant coefficients. In the same manner I find that

$$\sum \frac{\pi_r(e)}{K_r(b)} = \psi(u)M(e) \quad (24)$$

in which  $\psi(u)$  is a function of only  $u$  and  $M$  is a homogeneous polynomial of fourth degree in  $e$  with constant coefficients. Equation (20) becomes

$$\varphi(v)M(e) = \frac{1}{\psi(u)L(v)}.$$

Since the first member is independent of  $u$ , and the second of  $v$ , both are equal to the same constant  $C$

$$\pi(v) = \frac{C}{M(e)}; \quad \psi(u) = \frac{1}{CL(b)}. \quad (25)$$

From (18), (24), (25), and from (14), we obtain

$$P_*(e) = Q(b) \frac{M(e)}{CL(b)}, \quad R(u, v) = \frac{P(e)}{M(e)} \frac{\Lambda(b)}{Q(b)} \quad [\Lambda(b) = CL(b)] \quad (26)$$

in which all the polynomials are homogeneous, of fourth degree with constant coefficients. But  $R(0, v) = 1$ , and  $\Lambda(b)/Q(b)$  is a constant, if  $u = 0$ . Therefore  $P(e)/M(e) = k$  const., and this equation must be an identity, since the curve  $\Gamma_u = 0$  ( $x_i = e_i(v)$ ) does not belong to an algebraic surface. In the same manner one realizes that  $\Lambda(b)/M(b)$  is a constant. From (26)

we deduce consequently that  $R$  is a constant. And, since  $R(0, v) = R(u, 0) = R(0, 0) = 1$ , we deduce that

$$R(u, v) = 1.$$

Equation (8) becomes

$$\sum a_{ik}(v)b_k(u) = \sum c_{ih}(v)e_h(v).$$

By giving to  $u$  four particular values  $u = u_1, u_2, u_3, u_4$ , and solving these equations with respect to  $a_{i1}, a_{i2}, a_{i3}, a_{i4}$ , we obtain

$$a_{ij} = \sum_h \lambda_{ijh} e_h(v) \quad (\lambda_{ijh} = \text{const.})$$

and therefore

$$\sum_{k,h} \lambda_{ikh} b_k(u) e_h(v) = \sum c_{ir}(u) e_r(v).$$

As usual, this equation is an identity, and therefore

$$C_{ir} = \sum_k \lambda_{ikr} b_k(u).$$

The parametric equations of our surfaces are

$$x_i = \sum_{k,h} \lambda_{ikh} b_k(u) e_h(v)$$

with  $\lambda$  arbitrary constants.

The existence of these surfaces  $S$  was obvious *a priori*. It may perhaps be interesting to remark that, besides these surfaces, there is no other surface  $S$ , if neither the curve  $u = 0$ , nor the curve  $v = 0$  belongs to an algebraic surface. If it is not so, for instance, if  $S$  is a ruled surface and the line  $u = 0$  is a generatrix, the theorem can be completely false.

*GALACTIC AND EXTRAGALACTIC STUDIES, XIII. NOTE ON  
THE COMPARATIVE DIAMETERS OF SPHEROIDAL AND  
SPIRAL GALAXIES*

BY HARLOW SHAPLEY

HARVARD COLLEGE OBSERVATORY

Communicated April 13, 1942

From the photometric results presented in this communication it appears that contrary to some earlier indications the open spirals, with their supergiant stars and superficial structure of spiral arms, are not appreciably larger than spheroidal galaxies, which seem to be devoid of highly luminous stars and free of distributional irregularities. It follows that if in the course of time spiral arms appear in a flattened spheroidal system, they probably should not be treated as an extension or an expansion outward from the nuclear part of the galaxy, but as a development of structure well within the main body of the system. Or if, with the direction of evolution reversed, it be assumed that the spiral arms, supergiant stars and diffuse nebulosities of the spiral systems can eventually disappear into the structureless smooth form of the typical spheroidal galaxy, then that transformation also should be treated as an internal readjustment, and not as a contraction; for at any given considerable distance from the nucleus the amount of light (and probably of mass) is now found to be about the same for spheroidal and spiral systems.\* Apparently there is no important "growth" in dimensions along the sequence of forms.†

The spiral arms as observed are a phenomenon of only the inner half of an average galaxy. Moreover, less than twenty per cent of the light of a spiral galaxy is, on the average, in its spiral arms. The remainder is mostly in the commonly overlooked background in which the spires are embedded.

Since galaxies, like planetary atmospheres, probably fade out indefinitely, the extreme diameters are not very important in themselves (if the peripheral masses are negligible); but the dimensions out to a given light or mass density become very significant if evolutionary trends are under consideration and comparative sizes and gradients must be discussed.

1. A densitometric analysis of photographs of 123 bright northern galaxies was recently undertaken at the Harvard Observatory, chiefly to determine the density gradients and thereby to contribute to our fragmentary knowledge of the distribution of light and mass in large stellar systems of various types. Dealing comparatively with a large number of objects of all types, this work supplements the detailed analyses for individual galaxies by Reynolds, Hubble, Oort, Shirley and Redmond, Lind-

blad, Seyfert and others. The treatment of the gradients by Dr. Shirley Patterson will soon be published in the *Harvard Annals*.

A by-product of Miss Patterson's photometry has been the measurement of boundaries. From this material the quantitative intercomparison of the dimensions of spheroidal and spiral galaxies now becomes possible. The observational material is not extensive, especially for the spheroidal type; but it seems to be considerably the best now available for this particular problem, and is, in fact, about the first that seems to be precisely suited to the intercomparison of dimensions.

Although it is clear that the interpretation of the well-known sequence of types requires quantitative information of the sort now obtained on the relative dimensions, we should not overvalue the results for dynamical analysis, because our photographs measure only the distribution of light; the mass distribution may be otherwise. The ratio of mass to luminosity may, indeed, vary from type to type; almost certainly it varies from center to periphery in some of the galaxies.

2. From a general inspection of photographs, the conclusion has been easily reached in the past that spheroidal galaxies are on the average much smaller than the spirals, and that the more open spirals are largest of all. For instance, Hubble has tabulated the following results from reflector plates available at Mount Wilson:<sup>1</sup>

TYPE	DIAMETER	TYPE	DIAMETER
E0	0.58 kpc	Sa	1.84 kpc
E3	0.86	Sb	2.33
E7	1.47	Sc	2.92

He recognizes that these diameters refer to the "main bodies" of the galaxies. They show a marked and significant trend, with the average spiral having more than twice the linear diameter of the average spheroidal galaxy. In fact, the spherical E0 type, as measured by Hubble, compares more closely with the nucleus or inner disc of type Sc than with the five-times-larger spiral itself, and this circumstance has naturally suggested that spiral arms are dynamical expulsions or extensions from the nucleus, if there be assumed a developmental sequence in the direction E0-E7-Sa-Sc.

This same trend in the measured dimensions is evident in the Shapley-Ames catalog, where the diameters are also based on direct eyepiece measurement of reflector or refractor plates;<sup>2</sup> but when the diameters of the galaxies were more closely studied on a homogeneous series of long exposure Harvard plates,<sup>3</sup> the "growth" became less pronounced and the measured linear diameters of all galaxies were greatly increased over the previously used values. For example, a comparison of these Harvard measures of 1934 with Mount Wilson results shows that for forty-two



TABLE 1  
DIAMETERS OF 112 EXTERNAL GALAXIES

NGC	FSP TYPE	PG. MAG.	DIAM. IN kpc	NGC	FSP TYPE	PG. MAG.	DIAM. IN kpc
450	Sc2	12.6	5.0	4429	Sa	11.3	5.3
470	Sc2	12.7	3.6	4435	S0	11.7	4.4
474	E2.3	13.2	4.0	4442	SBa	11.5	4.1
488	Sa	11.8	4.4	4450	Sc1	11.1	4.7
628	Sc1	11.8	5.6	4459	E2.4	11.5	3.6
991	Sc2	12.7	5.4	4472	E1.0	10.1	7.0
1022	Sc2	12.0	2.7	4473	E4.1	11.4	4.6
1035	Sc1:	12.8	4.7	4486	E1.4	10.2	4.5
1042	Sc1	12.5	6.0	4490	Sc3	10.8	2.4
1052	E2.8	11.6	3.1	4501	Sc2	10.6	2.8
1084	Sc3	11.2	2.3	4503	S0	12.8	2.9
2403	Sc1	10.2	4.4	4517	(Sc)	11.6	8.4
2903	Sb1	10.3	4.0	4526	Sa	10.7	5.2
3031	Sb1	8.1	4.2	4527	Sc1	11.3	4.9
3115	E6.2	9.8	2.0	4535	Sc1	11.0	5.3
3184	Sc2	11.8	3.0	4536	Sc2	11.2	3.4
3198	Sc2	11.7	4.5	4548	Sc1	11.9	5.1
3338	Sc2	12.2	5.0	4550	Sa	12.8	2.9
3351	Sc1	11.5	5.4	4552	E0.0	11.7	4.8
3368	Sa	10.4	2.9	4559	Sc1	10.7	3.8
3556	Sc2	11.0	3.9	4564	S0	12.1	3.1
3623	Sc2	10.5	3.2	4565	Sb1	10.7	5.8
3627	Sc2	9.9	2.8	4567	Sc2	12.3	3.4
3628	(Sb)	11.3	6.7	4568	Sc2	12.2	4.0
3631	Sc3	11.8	2.2	4569	Sc1	11.3	13.6
3917	Sc3	12.8	5.0	4571	Sc1	12.8	3.8
3958	Sc2	11.5	3.2	4579	Sc1	10.8	11.5
3992	Sc2	11.2	4.2	4621	E2.7	11.8	4.2
4013	Sc1	12.7:	6.1	4631	Sc1	9.6	6.3
4036	Sa	11.9	4.0	4647	Sc2	12.7	4.9
4051	Sc1	11.7	4.4	4649	E1.2	11.2	5.7
4088	Sc3	11.2	2.3	4654	Sc2	11.7	4.7
4096	Sc1	12.2	6.0	4656	Sc2	11.3	3.8
4100	Sc2	11.9	3.2	4725	Sb1	10.8	4.8
4111	S0	11.6	3.9	4736	Sb1	9.0	1.9
4138	Sa	12.2	4.2	4826	Sc2	8.0:	2.8
4143	Sa	12.2	2.6	5005	Sc3	11.3	2.7
4144	Sc2	12.4	7.2	5033	Sb1	11.8	6.5
4157	Sc1	12.0	5.1	5055	Sb1	10.5	5.6
4192	Sc1	11.1	7.2	5194	Sc1	10.1	3.3
4214	Sc1	10.7	4.1	5371	Sc2	11.7	2.6
4216	Sc1	11.2	6.9	5377	Sa	12.8	5.9
4236	Sc1	11.3:	6.2	5457	Sc2	9.0	2.9
4242	Sc2	11.8	4.3	5474	Sc2	11.7	2.7
4244	Sc1	11.0	8.9	5631	Sa	12.5	3.1
4254	Sc2	10.5	4.5	5660	Sc3	12.3	2.7
4258	Sb1	10.2	6.8	5676	Sc3	11.9	3.4
4274	Sa	11.7	4.1	5689	SBa	12.6	3.5
4281	Sa	12.6	4.4	5866	Sa	11.5	4.5
4303	Sc1	10.4	5.8	5907	Sc	11.8	8.8
4321	Sc1	10.5	6.3	5985	Sc1	12.2	4.1
4371	SBa	12.1	4.6	6015	Sc3	12.1	3.2
4374	E2.1	10.7	5.7	6503	Sc3	11.4	4.2
4382	E2.7	10.3	7.0	6946	Sc1	11.1:	4.7
4406	E1.0	10.9	7.3	7217	Sc1	12.2	2.9
4414	Sb1	11.3	4.2	7479	Sc2	12.3	3.8

spheroidal galaxies the diameters are increased by an average of 3.2 times, but for eighty spirals only 1.6 times. Thus the large difference between

the two general classes was lessened, even when only visual estimates were employed on non-standardized photographs.

At the same time some preliminary measurements with a microdensitometer of a few of the same series of Harvard plates (three-hour exposures with the Bruce refractor) indicated that when using simple eyepiece measures we are far from exhausting the evidence on dimensions. Since some galaxies are common to the various dimensional surveys, we can show the advantage of densitometric measurement by determining mean ratios of the measures of diameters by the various methods.<sup>3</sup> We have the ratios

Microdensitometer to Mount Wilson estimates = 4.1 (9 objects)

Microdensitometer to Harvard 1934 estimates = 1.6 (14 objects)

Since this early densitometric work was based on photographs that were not precisely standardized, a study of the overall dimensions of galax-

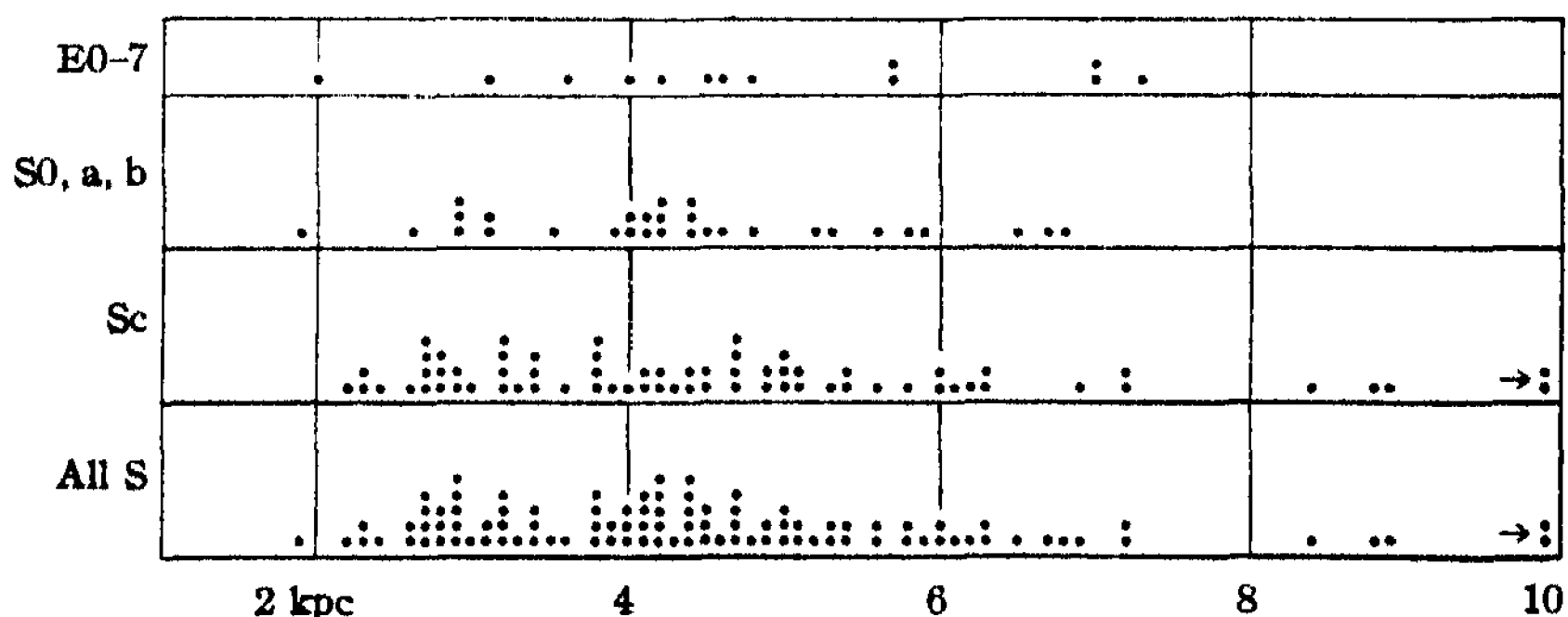


FIGURE 1

Distribution of diameters for 112 galaxies. Ordinates are numbers; abscissae are major diameters in kiloparsecs.

ies was then not carried further. The problem is now taken up again because suitable photographs have been made. The plates were exposed and developed under standard conditions by Miss Patterson. The 12-inch Metcalf refractor at Oak Ridge is especially suited to the work because of its large flat field and its suitable focal length. The plates have been systematically calibrated with a tube photometer, developed in a way that avoids Eberhard effect, and measured with a uniform technique. A description of the plate material and the methods involved will be described elsewhere.<sup>4</sup>

3. The linear diameters are given for 112 galaxies in table 1. Although they, too, have been analyzed photometrically, the nearby systems M 31,

32, 33, are omitted, as are also five irregular galaxies, and three classed as S without subdivision. The tabulated quantities indicate the extreme lengths of the measurable major axes. On a densitometer tracing the border of a galaxy is indicated by the clear rising of the density curve from the background of sky and plate fog. The important point in the present intercomparison is the strict comparability of the light intensity, for all types of galaxies, near the ends of the respective major axes.

The apparent magnitudes in table 1 are photoelectric (Stebbins and Whitford), when available, or photographic (*Harv. Ann.*, 88, No. 2), except that 8.1 is now adopted for NGC 3031. The linear diameters are based on the best values of the distances at present available.

4. In figure 1 the distribution of the diameters is graphically shown for the various subtypes. The smallest major diameters are about 2000 parsecs; the median and mean values for all 112 objects are more than double that value, and are practically the same for all types represented in the figure. The averages can be summarized as follows:

TYPE	NUMBER	MEDIAN DIAMETER, KPC	MEAN DIAMETER, KPC	MEAN ERROR
E0-E7	13	4.6	4.88	$\pm 0.44$
S0, a, b	30	4.2	4.35	$\pm 0.23$
Sc	69	4.3	4.67	$\pm 0.25$
All S	99	4.2	4.57	

In view of the dispersions and the uncertainties of measurement, we infer from these tabulated results that the galaxies of various types are of strictly comparable dimensions.

5. It could be argued that the large value of the average diameter of the spheroidal galaxies arises from the selection for densitometric work of the brightest and largest of that class. From various earlier studies it is known that the spheroidal and spiral systems are more or less evenly scattered throughout all observed magnitudes in clusters of galaxies as well as in open metagalactic space. In consequence the dispersions in absolute magnitude are much alike.

To test for evidence of selection in the densitometric study, let us consider all galaxies listed in the fairly homogeneous Shapley-Ames catalog between right ascensions  $12^h 0$  and  $13^h 0$  and declinations  $0^\circ$  and  $+20^\circ$ , thus including practically all the recognized members of the Virgo cluster. As shown below, the mean apparent magnitude of the thirty-six spheroidal galaxies is 11.98; for the eighty spirals it is 12.09. The magnitude difference of 0.11 should be compared with the difference of 0.51 in the last column of the following tabulation where only those Virgo galaxies are included that were analyzed with the densitometer.

TYPE IN <i>Harv. Ann.</i> , VOL. 88, NO. 2	ALL IN VIRGO REGION	ANALYZED BY DENSITOMETER
E0-E7	11 <sup>m</sup> .98 (36)	11 <sup>m</sup> .07 (11)
S, S0, a, b, c	12.09 (80)	11.58 (27)
Magnitude difference	0.11	0.51

The tabulated quantities are mean magnitudes, with numbers of galaxies in parentheses.

Obviously there has been some selection of the bright spheroidal systems for the densitometric work—a selection on the basis of apparent (and absolute) magnitude, and it amounts in the mean to  $\Delta m = 0^m.4$ . To correct the average diameter,  $D$ , for this selection we have

$$\Delta \log D = -0.2 \Delta m = -0.08$$

and to the mean and median diameters of the measured spheroidal galaxies should be applied the reduction factor 0.83. The results corrected for selection are, therefore,

Spheroidal: median diameter 3.8 kpc, mean diameter 4.05 kpc

Spiral: median diameter 4.2 kpc, mean diameter 4.57 kpc

This reduction still leaves spheroidal galaxies much larger than the part of the spirals occupied by recognizable spiral arms, and leaves us with the conclusion, mentioned in the introduction, that the two main types of external galaxies are of essentially the same linear dimensions.

\* It is, of course, not strictly necessary to assume that galaxies actually develop from one type to another. But a static universe (so far as galaxy evolution is concerned), or a very recent "creation," or a common unknown parent of all galaxy forms, are among the less comfortable of possible hypotheses.

† These conclusions are not incompatible with Hubble's recent important observations on the nature of the transition form S0 and on the first appearance (or vanishing) of disc and spiral structure in the spheroidal-spiral series. See also Randers, *Mt. Wilson Contr.*, No. 634, p. 282 (1940).

<sup>1</sup> *The Realm of the Nebulae*, p. 178 (1936).

<sup>2</sup> *Harv. Ann.*, 88, No. 2 (1932).

<sup>3</sup> *Ibid.*, 88, No. 4 (1934); also *Harv. Bull.*, 895, p. 22 ff. (1934).

<sup>4</sup> *Harv. Ann.*, 88, No. 7 (in press).

GALACTIC AND EXTRAGALACTIC STUDIES, XIV. ON THE  
MAGNITUDE DISPERSION IN THE PERIOD-LUMINOSITY  
RELATION

BY HARLOW SHAPLEY, VIRGINIA MCKIBBEN AND RICHARD A. CRAIG

HARVARD COLLEGE OBSERVATORY

Communicated April 13, 1942

The number of Cepheid variable stars in the Small Magellanic Cloud for which we now have periods and light curves has been increased to 564. The data on the dispersion of median magnitudes about the mean period-luminosity curve are therefore sufficiently abundant to merit preliminary examination. Although it is still impossible to disentangle the various causes of deviation, we are able to study the effect of doubling (superposed images), to reexamine the evidence for space absorption within the Cloud, and to test the fit of the adopted period-luminosity curve.

The new periods—about two hundred and fifty in number—which have been determined during the past year, will be published soon in a Circular of the Harvard College Observatory.

1. *Causes of Dispersion.*—The spread of the median magnitudes for the Cepheids in any compact stellar system can be attributed to some or, more probably, to all of the following six principal factors:

- (1) irregular space absorption in the system,
- (2) superposition of star images,
- (3) extent of the system in the line of sight,
- (4) inherent spread of luminosities for stars of the same period,
- (5) Eberhard effect and background haze,
- (6) observational and computational uncertainties, including magnitude-sequence irregularity.

The first factor throws the point for the affected star below (fainter than) the average period-luminosity curve; the second throws it above, and somewhat diminishes the range. The other four can work in both directions.

2. *The Period-Luminosity Curve.*—The present status of the relation of apparent photographic median magnitude to the logarithm of the period is exhibited in the usual form in figure 1, where the curve is that adopted earlier.<sup>1</sup> The curve was based on data for 307 stars but reasonably fits the 564 now plotted. The survey for variables with median magnitudes fainter than 16.8 is probably not complete because of faintness and crowding; the lower part of the curve was drawn with this probable deficiency in mind.

3. *The Scatter Diagram of Magnitude Residuals.*—The dispersion in the magnitudes is best shown by the scatter diagram in figure 2, where the

abscissae are logarithms of the period and the ordinates are deviations in terms of magnitude. The errors in the periods are relatively so small that we ignore them in this discussion and treat the period-luminosity curve as a regression relation. The factors listed in the first section above produce significant dispersions in magnitude only, except that number (4), inherent spread, could of course be associated with length of period as well as to luminosity.

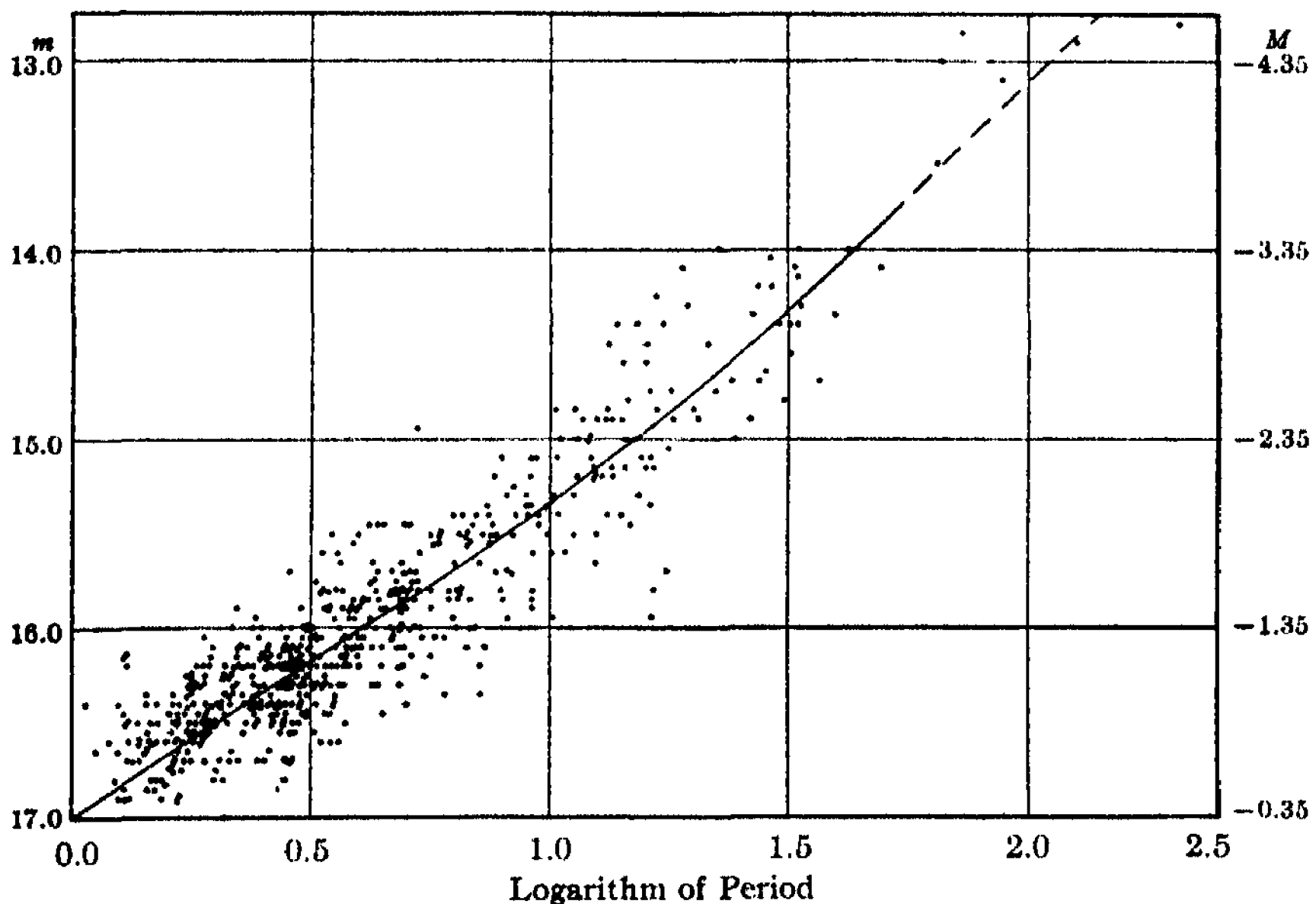


FIGURE 1

Period-luminosity observations for Small Magellanic Cloud. Ordinates are apparent magnitudes (left) and absolute magnitudes (right); abscissae are logarithms of the periods in days.

The plotted residuals in figure 2 have been computed for each variable from the relation

$$\Delta m = \dot{m}' - \dot{m}$$

where  $\dot{m}'$  is the observed median photographic magnitude and

$$\dot{m} = 17.07 - 1.74 \log P$$

This last relation is the straight line found, in the earlier discussion,<sup>1</sup> to represent the period-luminosity relation most satisfactorily for apparent median magnitudes in the Small Cloud. The adopted non-linear period-luminosity curve deviates only slightly, as shown in figure 4, from the

straight line; and in examining the relative dispersions for different periods, luminosities and positions in the Cloud, the linear relation suffices.

The spread in the median magnitudes is much the same throughout the whole range of periods (figures 2 and 4). At about sixteen days it is larger than average, but this is chiefly due to a few points far below the curve for which some of the magnitudes appear to suffer through intra-Cloud

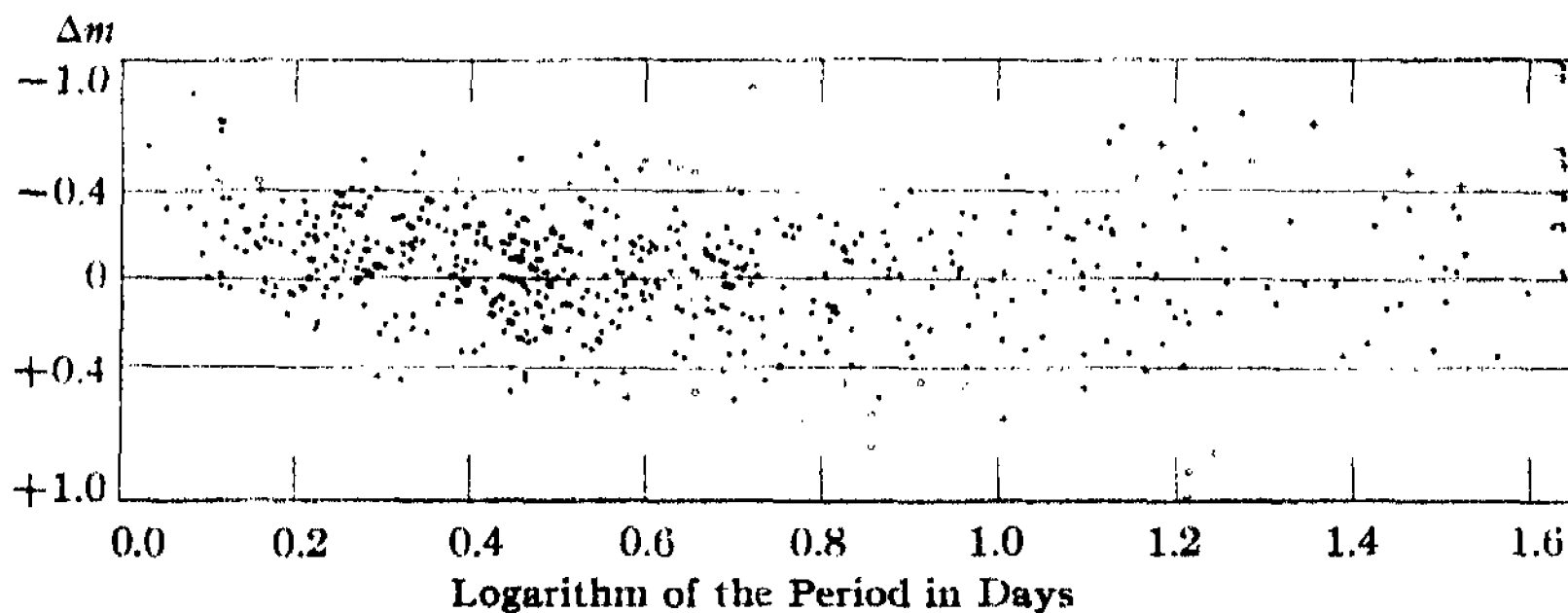


FIGURE 2

Magnitude residuals from the linear period-luminosity relation. For residuals greater than  $\pm 0.4$ , open circles refer to variables in the nucleus, crosses to variables in border regions.

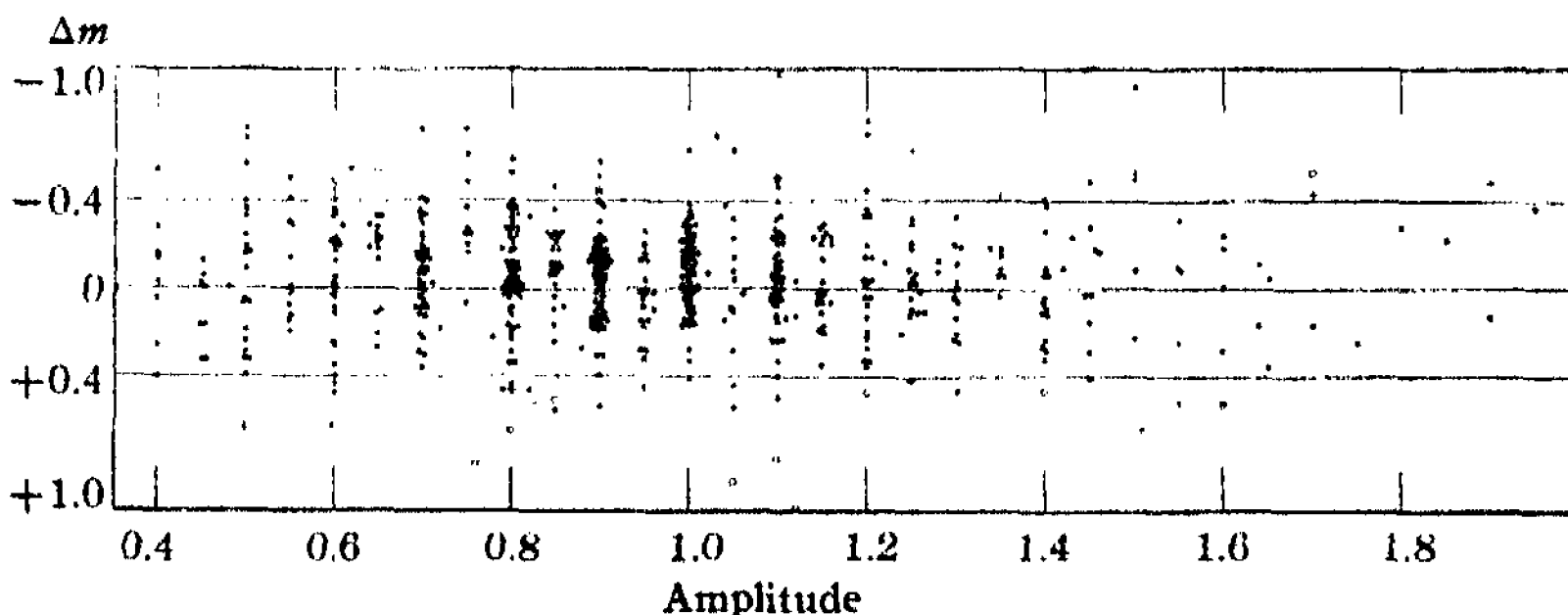


FIGURE 3

Magnitude residuals from the linear period-luminosity curve against amplitudes of variation. Open circles and crosses as for figure 2.

absorption; and it is somewhat less for periods of 1.5 days, almost certainly because of the incompleteness of our records for faint magnitudes.

4. *Dispersion and Amplitude.*—It is possible that the points lying well above the average curve (figures 1 and 2) refer in part to stars for which the measured brightness has been erroneously increased through unresolved "optical" companions having been measured along with the variables. If the deviations are due to such a cause, the measured amplitude  $A'$  of a

variable so involved should be somewhat less than the true amplitude  $A$ ; that is, the loss of amplitude through the effect of doubling is

$$\Delta A = (m_2 - m_1) - (m_2' - m_1')$$

where measured quantities are denoted by primes, and subscripts 1 and 2 refer to maximum and minimum light, respectively. If  $m_x$  is the magnitude of the superposed image, whether of one star or more, the true magnitude limits of the variable  $m_1$ ,  $m_2$ , are found by familiar formulae or tables,<sup>2</sup> since the corresponding luminosities are  $l_1 = l_1' - l_x$ ,  $l_2 = l_2' - l_x$ . The brighter the image  $x$ , the larger is  $\Delta A$ .

The affected stars should, of course, occur proportionately more frequently in the most populous parts of the Cloud.

To test the suggested explanation of the large negative deviations, we have plotted amplitudes against residuals in figure 3. The data for variables in the nuclear region are plotted as circles, and in the border regions as crosses, when the residuals exceed  $\pm 0.4$ . No correlation is obvious in this scatter diagram. The spread of amplitudes probably arises more from inherent causes than from superposed images.

The point is further examined in table 1, where mean amplitudes are given in specified intervals of deviation for the Cloud as a whole and for various regions of the Cloud. In what we call the nucleus, border and intermediate regions, the known variables have been studied with reasonable completeness;<sup>3</sup> the "other" variables are scattered throughout various other sections of the Cloud where the selection of variables for study has been somewhat superficial.

The similarity of the values of the mean amplitudes in the second column is striking. It means that in the Cloud as a whole the amplitudes are not related to the size and sign of the magnitude deviations from the period-luminosity curve. Equally striking is the similarity of mean amplitudes in nuclear, border and intermediate regions (last line of table 1). Location in the Cloud apparently does not systematically affect range of variation—either observationally, because of degree of crowding and nature of background, or physically, because of differences in gravitational field or other physical discriminants between nucleus and border of stellar systems.

Examining the various regions separately, however, we find that the results for the nucleus are strongly suggestive of a superposition effect. For all variables with  $\Delta m < -0.09$ , we find for the mean amplitudes

Nucleus	0 <sup>m</sup> .824,	36 stars
Borders	0.940	45 stars
Intermediate	0.930	63 stars
Other	0.990	103 stars
All	0.941	247 stars



TABLE 1  
MEAN AMPLITUDES IN RELATION TO CLOUD STRUCTURE AND TO MAGNITUDE DISPERSION

RESIDUALS	ALL CLOUD	NUMBER	NUCLEUS	NUMBER	INTERMEDIATE	NUMBER	BORDER	NUMBER	OTHER	NUMBER
< -0.4	0.96	46	0.91	8	0.89	7	1.16	12	0.89	19
-0.39 to -0.22	0.95	99	0.76	9	0.91	25	0.85	20	1.05	45
-0.21 to -0.10	0.93	102	0.82	19	0.96	31	0.87	13	0.97	39
-0.09 to +0.01	0.96	97	0.92	15	0.95	28	0.99	21	0.98	33
+0.02 to +0.14	0.98	102	1.03	23	0.94	27	1.01	20	0.95	32
+0.15 to +0.39	0.96	86	0.98	13	0.93	23	0.94	26	1.01	24
> +0.4	0.96	30	1.05	10	0.86	6	0.83	9	1.15	5
All	0.956	562	0.93	97	0.93	147	0.95	121	0.99	197

and compare these values with the mean amplitude,  $0^m.968$  (315), for all the other variables,  $\Delta m > -0.9$ , which presumably are not affected by doubling. Thus we have

$$\overline{\Delta A} = 0^m.968 - 0^m.824 = 0^m.144$$

for the nucleus (other regions appear unaffected). If this average difference in amplitude were assumed to be wholly produced by superposed images, the corresponding average deviation would need to be about  $-0^m.25$ , with the contributing star image,  $m_x$ , a magnitude and a half fainter than the variable at minimum. In the crowded star fields such superpositions seem entirely credible.

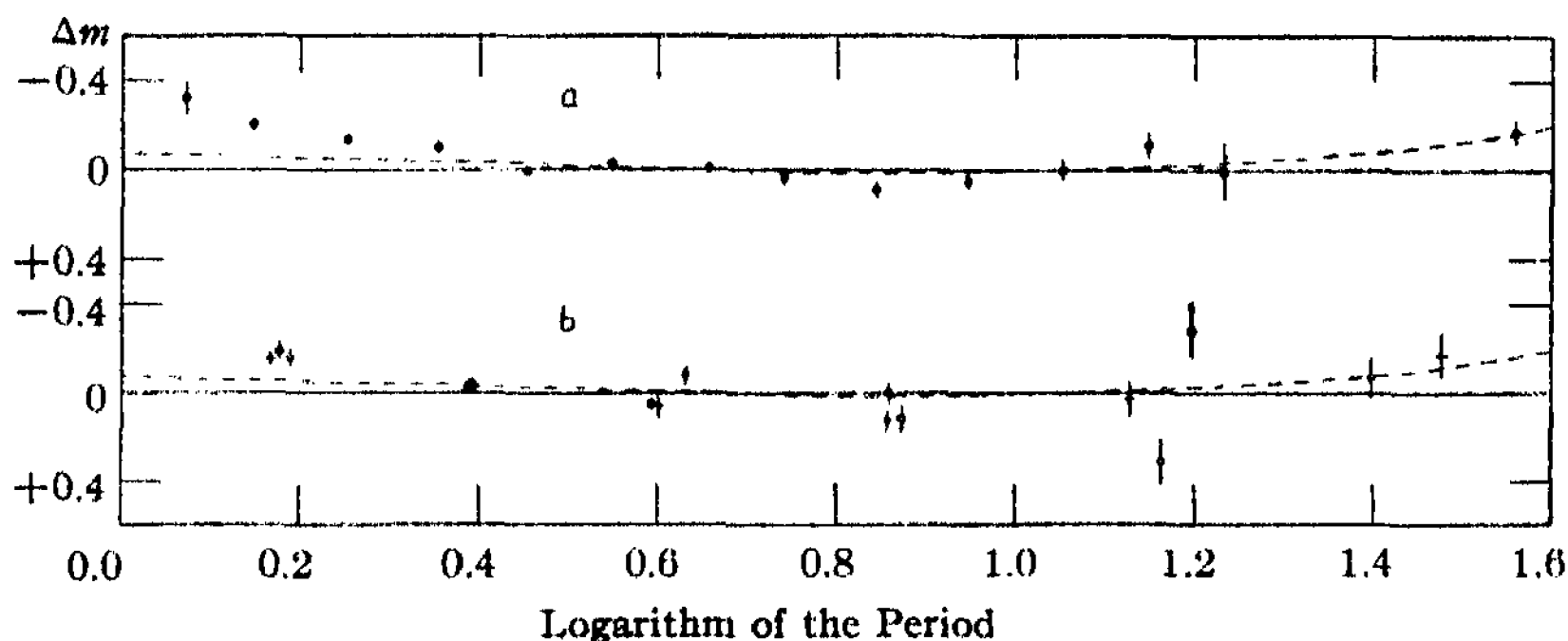


FIGURE 4

(a) Mean values for the material plotted in figure 2. Broken line represents adopted non-linear period-luminosity curve. Vertical lines through mean points indicate mean errors.

(b) Similar to (a) for selected regions of the Cloud. Circles refer to the nucleus, crosses to the border regions, dots to intermediate regions. The low point at  $\log P = 1.164$  includes some nuclear variables that are dimmed by obvious obscuration.

Since for stars of any given period there is probably a true dispersion in median luminosity, there may also be a dependence of amplitude on the median luminosity—a dependence which would work systematically to cancel or enhance any contribution to the scatter from superposed images. But unless such a dependence is operating in the sense that the more luminous variables (for a given period) have on the average greater amplitudes, and of just the amount necessary to cancel most of the effect from superposed images, we must conclude from the data of table 1 that except in the nuclear regions the contribution of doubling to the general dispersion in median magnitude is not important. This conclusion is supported by the appearance of the star fields, since the chances of trouble with superposed images, except for the fainter variables in the denser star fields of the

Cloud, seem to be slight for plates made with the Bruce refractor or the 60-inch reflector.

5. *Test of the Period-Luminosity Curve.*—In figure 4 (a) mean deviations are plotted against logarithms of the periods, for intervals of 0.1 in the logarithm, and are given also in table 2, along with the mean error of the mean deviation and the corresponding mean logarithm of the period. The mean errors are indicated in the figure by vertical lines through the points. The adopted non-linear period-luminosity curve has also been drawn in the figure. Although throughout its course it does not appear to be the best possible fit for this larger amount of material, it so nearly represents the most heavily weighted portion of the plot that a further revision at this time seems unnecessary and inadvisable. The lack of fit for the shortest periods and our reason for ignoring it have been mentioned above.

TABLE 2  
RELATION OF MAGNITUDE DISPERSION TO LENGTH OF PERIOD

INTERVAL OF LOG OF PERIOD	MEAN LOG OF PERIOD	MEAN RESIDUAL	MEAN ERROR	NUMBER
0.0-0.1	0.071	-0 <sup>m</sup> .32	±0.08	5
0.1-0.2	0.148	-0.20	±0.04	37
0.2-0.3	0.252	-0.14	±0.02	67
0.3-0.4	0.355	-0.10	±0.03	59
0.4-0.5	0.454	0.00	±0.02	90
0.5-0.6	0.549	-0.03	±0.03	69
0.6-0.7	0.656	-0.01	±0.03	53
0.7-0.8	0.740	+0.03	±0.05	34
0.8-0.9	0.844	+0.09	±0.05	33
0.9-1.0	0.947	+0.05	±0.05	24
1.0-1.1	1.051	0.00	±0.06	21
1.1-1.2	1.149	-0.11	±0.07	23
1.2-1.3	1.233	+0.01	±0.14	16
> 1.3	1.559	-0.17	±0.06	33

6. *Space Absorption Within the Small Cloud.*—In table 3 we have tabulated mean magnitude deviations for intervals of 0.25 in the logarithm of the period for the nuclear, intermediate and border regions, separately. The numerous variables that do not lie in any one of these regions, where the survey has approached completeness, have not been included in this tabulation or in the corresponding figure 4 (b). Only 365 variables, rather than the 564 of figures 1, 2 and 4 (a), are therefore involved. Because of the small numbers of stars represented, the mean errors of some of the means are rather large. The adopted luminosity curve has been included, as in figure 4 (a).

If there were appreciable space absorption in the nuclear part of the Cloud, the apparent median magnitudes for stars in that region should be fainter than normal, and the plot in figure 4 (b) should reveal this system-

atic effect. But to the accuracy permitted by the mean errors of the plotted points, we can say that space absorption in the Small Magellanic Cloud is inappreciable. This conclusion does not mean that the Small

TABLE 3  
PERIOD, MAGNITUDE DISPERSION AND CLOUD STRUCTURE

INTERVAL OF LOG OF PERIOD	MEAN LOG OF PERIOD	MEAN RESIDUAL	MEAN ERROR	NUMBER
<i>Nucleus</i>				
0.00-0.25	0.190	-0.16	$\pm 0.05$	12
0.25-0.50	0.386	-0.02	$\pm 0.03$	23
0.50-0.75	0.631	-0.08	$\pm 0.06$	22
0.75-1.00	0.874	+0.12	$\pm 0.07$	21
1.00-1.25	1.164	+0.31	$\pm 0.11$	12
> 1.25	1.398	-0.07	$\pm 0.10$	7
<i>Intermediate</i>				
0.00-0.25	0.178	-0.19	$\pm 0.05$	21
0.25-0.50	0.391	-0.04	$\pm 0.03$	65
0.50-0.75	0.593	+0.05	$\pm 0.03$	38
0.75-1.00	0.860	0.00	$\pm 0.06$	16
> 1.00	1.198	-0.28	$\pm 0.14$	7
<i>Border</i>				
0.00-0.25	0.168	-0.15	$\pm 0.03$	21
0.25-0.50	0.394	-0.03	$\pm 0.04$	39
0.50-0.75	0.601	+0.06	$\pm 0.07$	21
0.75-1.00	0.858	+0.13	$\pm 0.06$	15
1.00-1.25	1.129	+0.03	$\pm 0.09$	16
> 1.25	1.476	-0.17	$\pm 0.11$	10

Cloud is wholly free of absorbing material; a few bright and dark patches have been observed. Nor does it refer to the space absorption between the observer and the Small Cloud. That absorption, which is believed to amount to three-tenths of a magnitude in photographic light, has been indirectly detected through the scarcity on Harvard long-exposure plates of remote external galaxies in the direction of the Cloud.

<sup>1</sup> Eighth paper of this series, these PROCEEDINGS, **26**, 543-544 (1940).

<sup>2</sup> For example, *Harv. Ann.*, **33**, 287 (1900).

<sup>3</sup> See fifth paper of this series, these PROCEEDINGS, **26**, 110 (1940).

*GALACTIC AND EXTRAGALACTIC STUDIES, XV. ON THE  
DISTRIBUTION OF PERIODS FOR 343 CEPHEIDS IN THE SMALL  
MAGELLANIC CLOUD*

BY HARLOW SHAPLEY AND VIRGINIA MCKIBBEN

HARVARD COLLEGE OBSERVATORY

Communicated April 13, 1942

In the present communication we treat further the distribution of periods of classical Cepheids, a subject that has already been considered in the fifth paper of this series<sup>1</sup> for the Magellanic Clouds and in the eleventh<sup>2</sup> for the nucleus of our own Galaxy.

In both of the earlier discussions it was found that the conventional period-frequency curve, which is based on the galactic Cepheids that are chiefly in the solar neighborhood, is apparently of only local applicability. Not only does it fail to represent the distribution of periods found in the Magellanic Clouds, but it fails also at the galactic nucleus and elsewhere. Moreover, because of the dependence of period on absolute magnitude, and the consequent favoring of the longer periods in any survey based on apparent magnitude, the solar neighborhood cannot be impartially explored; and therefore the conventional curve of period frequency is not accurately representative even of the space around the sun. An appropriate correction for this absolute magnitude selection changes the local curve toward a resemblance to the frequency curve for the Small Cloud, but only partly cancels the observed differences.

We also give in the present paper a preliminary determination of the distance of the "wing" of the Small Magellanic Cloud, a comment on the classical Cepheids in the globular clusters, and a sample of the light curves of the "16-day" Cepheids.

1. *The Distribution of Periods.*—Of the 564 classical Cepheids that have now been studied in the Small Cloud, and for which periods and magnitudes have been discussed in the preceding paper of the series, 343 lie within the nuclear, intermediate and border regions, where the search for and measurement of the variables have been reasonably complete. The other 221 variables are outside these regions, in parts of the Cloud where the survey has been incomplete and where a preferential selection of the brighter stars and the longer periods has no doubt operated. These outside variables are obviously not suitable for a study of the period frequency; in fact, they give, in their incomplete representation, a distribution curve that rather closely resembles the curve for the 288 galactic Cepheids.<sup>1</sup>

The positions in the Cloud of the chosen nuclear, intermediate and border regions are the same as specified in the earlier paper, but all have been enlarged. It is believed that by using only the results in these

thoroughly worked regions we are getting the most nearly correct picture of the distribution of the periods of classical Cepheids in this particular galaxy; and perhaps it is the correct picture of the normal or average behavior of Cepheids in our Galaxy, when it is considered as a whole.

TABLE 1  
DISTRIBUTION OF PERIODS OF CEPHEIDS IN SPECIAL  
SECTIONS OF THE SMALL MAGELLANIC CLOUD

REGION	NUMBER	PERCENTAGE OF VARIABLES WITH PERIOD				
		<2 <sup>d</sup> %	2 <sup>d</sup> TO 10 <sup>d</sup> %	>10 <sup>d</sup> %	14 <sup>d</sup> TO 17 <sup>d</sup> %	>20 <sup>d</sup> %
Nucleus	74	10.8	63.5	25.7	10.8	8.1
Core	24	8.3	50.0	41.7	16.7	16.7
Intermediate	147	21.8	73.5	4.8	0.7	0.7
Borders	122	23.8	54.9	21.3	5.7	7.4
All	564	19.3	64.2	16.5	3.7	5.9

Peculiar distributions occur in special localities, such as the galactic nucleus and globular clusters, or at great distances above and below the galactic plane, where with rare exceptions Cepheids appear only in the form of cluster-type variables.

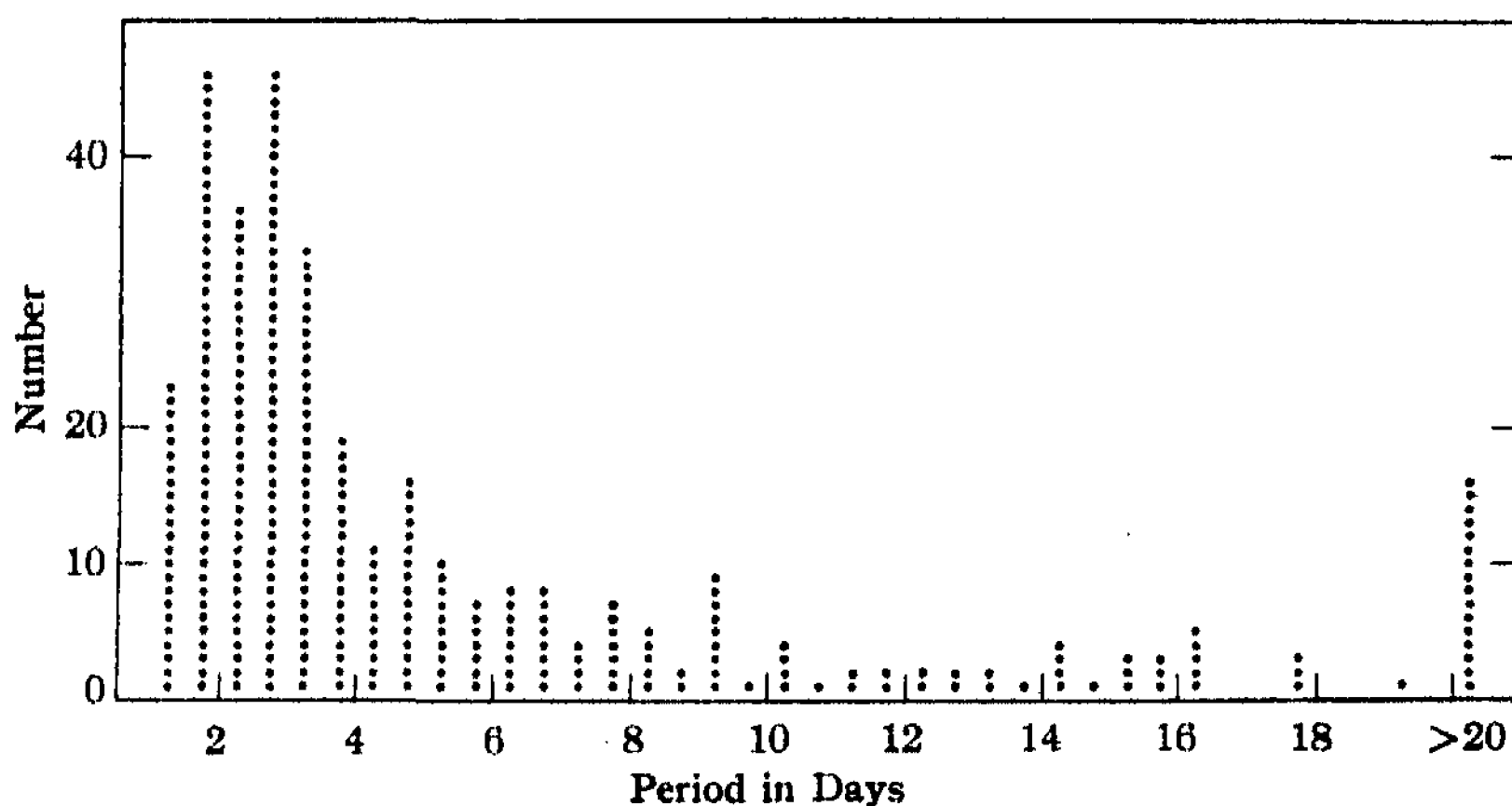


FIGURE 1

Period distribution of Cepheids in the Small Magellanic Cloud.

The material from the three regions is brought together in figure 1, which shows the number of variables for each half-day interval in the period. Nearly seventy per cent of the variables have periods less than 3.5 days, in striking contrast to only twelve per cent in our galactic system. The maximum of the frequency curve in the Small Cloud is close to two days; in our own system, 4.5 days. There are no certain secondary

maxima in the curves for the Cloud, although we note a considerable number of periods between fourteen and eighteen days. Periods of this length are also rather common among galactic Cepheids, and the unusual group in the direction of the galactic center is mentioned below.

In table 1 some data on the frequency of periods in the different regions of the Cloud are assembled. In making this tabulation twenty-three periods based on the study of plates made with the 60-inch reflector have been withdrawn from the material available for the nucleus because similar work with the reflector has not yet been completed for the borders and the intermediate regions. The larger scale of the reflector plates and their fainter magnitude limits permit the discovery of variables that have been overlooked on the Bruce plates.

Although the reflector plates reach well below the limit where cluster-type variables should appear, so far not one has been definitely found and

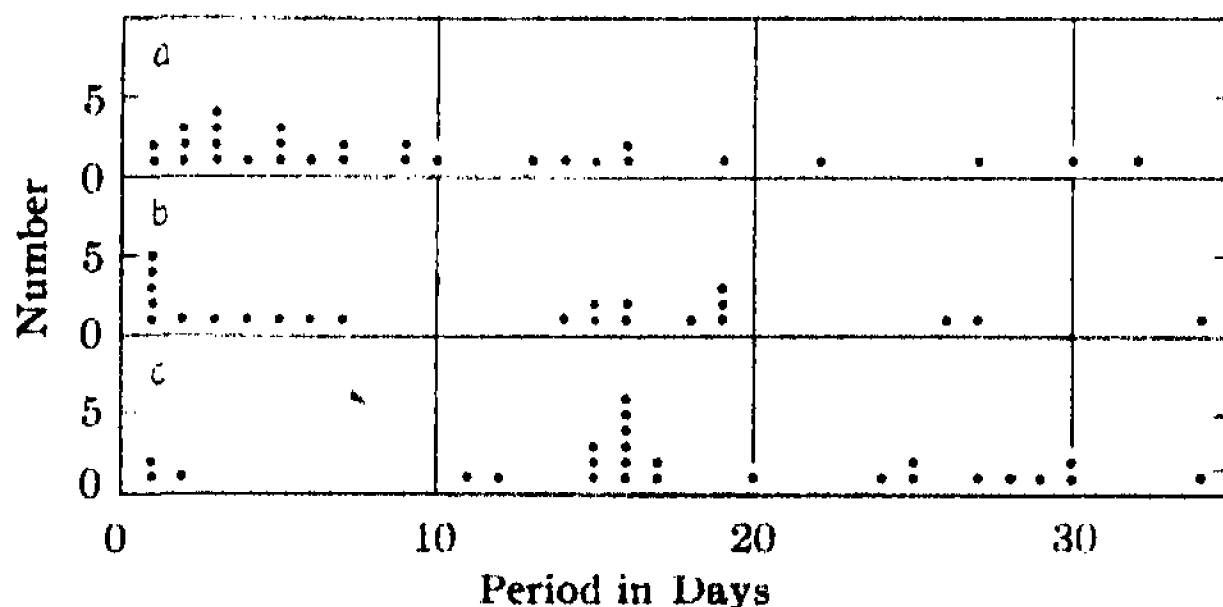


FIGURE 2

Period distribution of Cepheids in: (a) Core of Small Magellanic Cloud; (b) Globular clusters; (c) Galactic nucleus (longest period is 69.05 days).

assigned to the Cloud. It appears that if cluster variables exist in the Small Magellanic Cloud they are very scarce in comparison to classical Cepheids.

A few cluster-type variables are found on the photographs of the Cloud, but they are all brighter than the seventeenth magnitude and are not more abundant than would be expected for the foreground. Similarly a number of eclipsing stars, which probably are in the foreground rather than giant binaries in the Cloud, have been found and studied. It is difficult to be as certain concerning the non-membership of the several long-period variables. If they are in the Cloud they are of the supergiant type such as recorded for 47 Tucanae. Some of them certainly are foreground stars—typical long-period variables of our own galactic system.

2. *Anomalous Distributions of Period.*—The densest portion of the nucleus of the Small Magellanic Cloud has been called the “core.” Its

variables are listed separately in table 1. The great contrast in the distribution of periods between this core and the border regions was also noted in the earlier discussion. The cylindrical core extends, of course, from the near border to the far border, and the true contrast between the actual center of the Cloud and the border is therefore unattainable; we cannot tell which of the core or nuclear variables are really near the center,

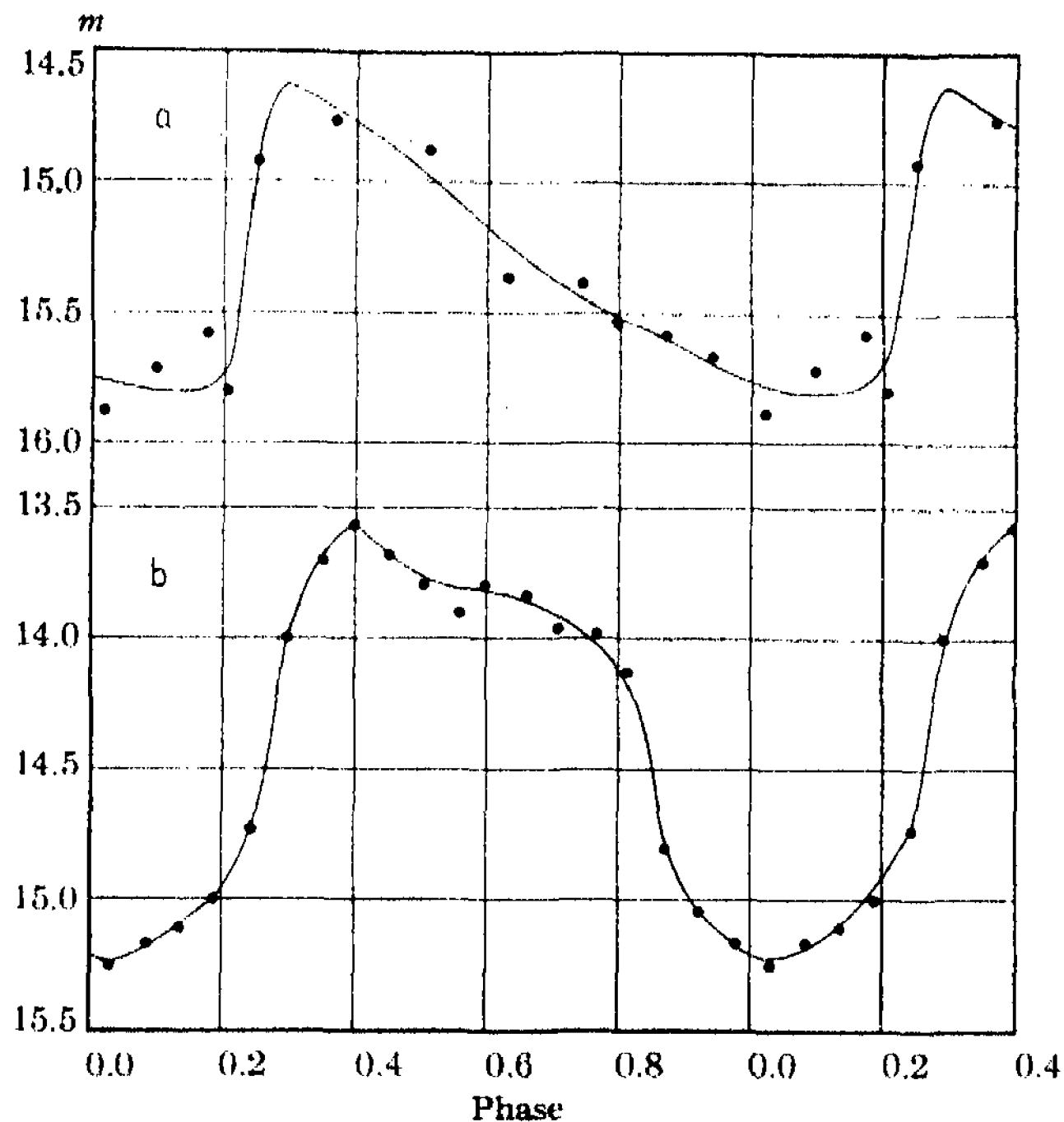


FIGURE 3

Light curves of two 16-day Cepheids: (a) HV 1333 and (b) HV 10495. Ordinates are photographic magnitudes; abscissae are fractions of the periods.

which far out in the cylinder. Consequently the contrast observed is a minimum.

The distribution for the core is shown graphically in figure 2, and in the same diagram is given the distribution for the twenty-three classical Cepheids in globular clusters,<sup>8</sup> and for the twenty-seven in the region of the galactic center. The data for the nucleus of the Galaxy are taken from the first table of the eleventh paper of this series, after removing five stars that are known from photometric data to be in the solar neighborhood and far this side of the center of the system. A comparison of figures 1 and 2



emphasizes the peculiar association of at least some of the long period Cepheids with the densest star fields; or perhaps the more striking point is the almost complete absence from these crowded regions of the average sort of Cepheid, with a period from two to seven days.

3. *Light Curves for the Sixteen-Day Cepheids.*—For convenience we call the Cepheids with periods between fourteen and seventeen days the "16-day" Cepheids. In the Small Magellanic Cloud these stars all appear to have typical Cepheid light curves, such as that suggested in figure 3 (a) for HV 1333. But about one-half of those in the globular clusters and at the galactic center have light curves much like that shown in figure 3 (b) for HV 10495. This uncommon light curve is like that of W Virginis,<sup>4</sup> a variable of period 17.27 days, which is, incidentally, in a very low density region far from the galactic plane. Much careful photometry and spectroscopy should be done on these 16-day Cepheids because of their possible significance in the study of the energy sources in supergiant stars.

4. *The Distance of the Small Cloud's Wing.*—A faint extension of the Small Magellanic Cloud in the general direction of the Large Cloud has been reported in earlier papers.<sup>5</sup> Whether this wing is actually a part of the Small Cloud cannot be told from general appearances, but it is now possible to apply the Cepheid test. The following four variables fall in the area of the wing:

NAME OF VARIABLE	PERIOD IN DAYS	MEDIAN MAGNITUDE	AMPLITUDE OF VARIATION
HV 2233	15.172	14 <sup>m</sup> .4	1 <sup>m</sup> .25
VMcK 14	16.2	15.35	0.9
EHB 37	21.4	14.5	1.4
HV 865	33.3	14.0	1.35:

The means of the median magnitudes and the logarithms of the period are  $14.56 \pm 0.30$  and  $1.31 \pm 0.08$ , respectively, and the corresponding distance modulus is

$$m - M = 17.18 \pm 0.34 \text{ (m.e.)}$$

This value should be compared with the distance moduli 17.10 and 17.35 for the Large Cloud and Small Cloud, respectively. To the accuracy permitted by the scanty data, we can conclude that the wing is probably a part of the Small Magellanic Cloud. Available photographs indicate that it does not have a dense background of faint stars, but there are a number of faint star clusters within its bounds, and some patches of diffuse nebulosity.

<sup>1</sup> These PROCEEDINGS, 26, 105–115 (1940); *Harvard Reprint* 192.

<sup>2</sup> Op. cit., 26, 681–688 (1940); *Harvard Reprint* 214.

<sup>3</sup> Miss Sawyer, *Publ. David Dunlap Obs.*, 1, No. 4 (1939).

<sup>4</sup> Gaposchkin, *Harv. Bull.*, 906, 8 (1937).

<sup>5</sup> *Harv. Bull.*, 914, 8 (1940); *The Telescope*, 8, 15 (1941).

## CLOSE LINKAGE BETWEEN MUTATIONS WITH SIMILAR EFFECTS

BY L. C. DUNN\* AND ERNST CASPARI†

COLUMBIA UNIVERSITY AND LAFAYETTE COLLEGE

Communicated April 2, 1942

About a dozen different mutations which affect the tail and other parts of the axial skeleton of the house mouse have been reported by various observers. Five of these appear to form a group of which the members are related to each other, first, by producing similar effects on the tail and, second, by showing little or no crossing-over with each other. The genetic relations among these five have now been tested with the result that three of them behave as though unilocal, while two others show less than 3% of recombination with the first locus or with each other. It appears thus that three adjacent loci have each produced mutations of similar kinds. This clustering of five similar mutations indicates a relationship between the location and the morphological effect of mutations which deserves careful study. A summary of the crossing-over data involving the five mutations is given herewith to be followed by a fuller discussion elsewhere.

The mutations involved are listed below, with references to the original or most complete description of their effects.

*T*, dominant Brachyury; tail short or absent; varies in different stocks; occasionally has fused or reduced vertebrae (Dobrovolskaia-Zawadskaia, 1934); *TT* is lethal (Chesley, 1935).

*Fu* (= *T'*) dominant Fused; fused and reduced vertebrae; moderately shortened; *Fu Fu* is viable (Reed, 1937).

*Ki*, dominant Kink; effect like *Fu*; *Ki Ki* is lethal (Caspary and David, 1940).

*t<sup>0</sup>*, recessive lethal; *Tt<sup>0</sup>* is tailless (Chesley and Dunn, 1936).

*t<sup>1</sup>*, recessive lethal; *Tt<sup>1</sup>* is tailless (Dunn and Gluecksohn-Schoenheimer, 1938).

(*t<sup>2</sup>*, recessive lethal like *t<sup>1</sup>*; *Tt<sup>2</sup>* is tailless, *t<sup>1</sup>t<sup>2</sup>* is lethal (Dunn, 1939) discarded; and not used in further work.)

All combinations of the five mutations have now been observed and a study of the phenotypic effects permits the following statements. *T*, *Fu* and *Ki* are clearly dominant in all combinations and have effects on the tail which show overlapping variability so that the three mutants cannot be reliably distinguished when heterozygous. The three dominants differ in their reactions with *t<sup>0</sup>* and *t<sup>1</sup>*, since *Tt<sup>0</sup>* and *Tt<sup>1</sup>* are tailless while *Fu t<sup>0</sup>*, *Fu t<sup>1</sup>*, *Ki t<sup>0</sup>*, and *Ki t<sup>1</sup>* all have deformed tails of the same sort, that is, like Brachy or Fused. On this basis, we might place *Fu* and *Ki* in one group, and *T*, *t<sup>0</sup>* and *t<sup>1</sup>* in another. Phenotypically there is further justification for such a grouping since animals with either *Fu* or *Ki* often show waltzing or choreic

behavior whereas these characters are not parts of the phenotypic effect of  $T$ ,  $t^0$  or  $t^1$ .  $T$ ,  $Fu$  and  $Ki$  produce similar phenotypes in combination with each other. Combinations of  $T$  with  $Fu$ ,  $T$  with  $Ki$  and  $Fu$  with  $Ki$  may have somewhat more extreme effects on the tail than the separate mutations, but the material is too variable to permit exact statements. On the basis of mutual interaction effects we might then recognize two groups: (1)  $Fu$  and  $Ki$ , (2)  $T$ ,  $t^1$  and  $t^0$ .

In genetic behavior no such separation into groups has heretofore been possible.  $T$ ,  $t^0$  and  $t^1$  have shown no crossing-over among themselves in over 6000 observations (Dunn unpublished). Reed (1937) did not observe any certain crossovers between  $Fu$  and  $T$  among 240 offspring from matings of  $\frac{Fu +}{+ T}$  by normal. The tests were complicated by the fact that  $Fu$  frequently failed to manifest itself, necessitating progeny tests of all suspected crossovers.

We have carried out new experiments designed to detect crossing-over between  $Fu$  and  $T$  (data chiefly from Dunn),  $Ki$  and  $T$ ,  $Ki$  and  $Fu$  (data chiefly from Caspari); between  $Ki$ ,  $t^0$  and  $t^1$  and  $Fu$ ,  $t^0$  and  $t^1$  (Dunn).

The results for  $Fu$  and  $T$  are shown in table 1.

PARENTS	OFFSPRING				
	ABNORMAL TAIL $T$ OR $Fu$	TAILLESS	NORMAL TAIL	NORMALS TESTED $Fu$ (OVERLAPS)	NORMALS TESTED $++$ (RECOMBI- NATIONS)
$++ \times \frac{Fu +}{+ T}$	356	4	13	6	4
$\frac{Fu +}{+ T} \times ++$	30	..	2	..	..
Total	386	4	15	6	4

The normal-tailed offspring arise either from recombination between  $Fu$  and  $T$  or from failure of one of these dominants to express itself in the offspring. Ten of the normal offspring have been sufficiently tested by crossing with normals. Six of them have produced both Fused and normal progeny showing them to be  $Fu +$ ; four of them, however, have produced, respectively, 37, 37, 33, 26 normal-tailed offspring, proving them to be  $++$ . Out of a total of 400 offspring (386  $Fu$  or  $T$ , 4 tailless and 10 tested normals) a minimum of four recombinations of the  $++$  type occurred.† If the test cross was  $\frac{Fu +}{+ T} \times ++$ , a contrary type of crossover,  $\frac{Fu T}{++}$ , is to be expected in equal numbers. These have not yet been detected, although it is likely that they are to be sought among the tailless animals which are weak and often sterile. The amount of recombination detected therefore repre-

sents half of the expected amount. The amount of recombination between *Fu* and *T* may thus be estimated roughly at about 2.0% (8/400).

From *inter se* matings of animals known to be  $\frac{Fu +}{+ t^0}$ , 123 Fused and 10 normal offspring have been obtained. Tests of the normals have not been completed. Three of them are known to be  $\frac{Fu +}{+ t^0}$  (overlaps).

We have carried out similar tests with Kink and Brachy. At Columbia, four males and two females heterozygous for these two mutations  $\left(\frac{Ki +}{+ T}\right)$  were crossed with normal animals ( $++$ ) and have given 408 offspring of which 406 were either Kink or Brachy, one male had a normal tail and one was tailless. The normal-tailed male when tested by normal-tailed females gave 39 offspring, all normal and was thus a  $++$  recombination (or a mutant).

The tailless male when tested by normal-tailed females gave 18 *tailless*, 21 normal and 10 Kink or Brachy. This suggested that this exceptional male was  $\frac{Ki T}{++}$ , that is, a recombination in which *Ki* and *T* were in the same chromosome. This was proved by mating him to Brachy females ( $T +$ ) and dissecting the mothers on the eleventh day of pregnancy. Typical *TT* homozygotes (cf. Chesley, 1935) were found. We are grateful to Dr. Gluecksohn-Schoenheimer for making these observations. Some of the abnormal-tailed offspring from matings of this male by normal females showed the waltzing behavior and deafness which occurs with *Ki* but never with *T* alone. These facts show him to contain *T* and most probably *Ki* in the arrangement  $\frac{T Ki}{++}$ . The fact that he was tailless (the only tailless animal out of 785 from similar ancestry) suggests that *Ki* and *T* may produce a greater interaction when both are present on the same chromosome.

The  $\frac{Ki +}{+ T}$  males tested by  $\frac{+ t^0}{+ t^1}$  females have given 192 with abnormal tails (98 *Ki*, 94 tailless  $Tt^0$  or  $Tt^1$ ) and 8 with normal tails. Three of the latter have proved to be  $+t^1$  or  $+t^0$  and are thus due to recombination, while three proved to carry *Ki*.

At Lafayette the result of crossing heterozygous females by normal males was 124 Kink or Brachy and 8 normal. The reciprocal cross yielded 245 Kink or Brachy and 8 normal. Of the 8 normals tested, 6 proved to be  $++$ . The total number of  $++$  recombinants is thus 10. Assuming these to represent half the total, the estimated frequency of recombination between loci *T* and *Ki* is about 2.0% (24/986).

The data are summarized in table 2.

TABLE 2

PARENTS	OFFSPRING				
	ABNORMAL TAIL ( <i>T</i> OR <i>Ki</i> )	NORMAL TAIL	NORMALS TESTED <i>Ki</i> (OVERLAPS)	NORMALS TESTED ++ (RECOMBI- NATIONS)	TAILLESS TESTED $\frac{Ki}{+} \frac{T}{+}$
$++X\frac{Ki}{+} \frac{+}{T}$ (C)	381	1	..	1*	1
$\frac{Ki}{+} \frac{+}{T}X++$ (C)	25	..	..	..	..
$++X\frac{Ki}{+} \frac{+}{T}$ (L)	245	8	2	3	..
$\frac{Ki}{+} \frac{+}{T}X++$ (L)	124	8	2	5	..
$\frac{+}{+} \frac{+}{+}X\frac{Ki}{+} \frac{+}{T}$ (C)	192	9	3	3†	..
Totals	967	26	7	12	1

(C). At Columbia.

(L). At Lafayette.

\* One male tested by ++ gave 39 normal progeny.

† Each animal tested by ++ gave 25 or more normal progeny. One male tested by ++ gave 30 normal and by *T*+ gave 12 normal, 1 Brachy, 8 tailless, hence is  $+t^1$  or  $+t^0$ . One male tested by  $++$  gave 43 normal and by *T*+ gave 12 normal, 9 Brachy, and 5 tailless, hence is  $+t^0$  or  $+t^1$ ; one male tested by normal, gave 25 normal and by *T*+ gave 8 normal, 6 Brachy, 9 tailless.

From matings of  $\frac{Ki}{+} \frac{+}{t^1}X\frac{Ki}{+} \frac{+}{t^1}$  93 offspring have been recorded of which 91 were Kink and 2 were normal. The latter have not yet been tested. Because of the close linkage between the two lethals *Ki* and *t*<sup>1</sup>, the stock containing both may be maintained as a balanced lethal stock by eliminating the rare normals which arise by recombination.

It has also been possible to test whether *Ki* and *Fu* are mutations at the same or different loci. Animals of genotype  $\frac{Fu}{+} \frac{+}{Ki}$  (having the appearance of Fused or Kink) crossed with normal-tailed testers gave 172 *Fu* or *Ki* and 21 normals. Of 11 of the latter, one when tested gave only normal-tailed offspring (61 normal test offspring and 1 with slightly abnormal tail). Another one may be a recombinant, too, having had 29 normal progeny only. The other 9 are overlaps.

These results appear to us to establish the fact that *T*, *Ki* and *Fu* are mutations at three separate loci. The relative recombination frequencies and the order of the loci cannot be established from the present limited data in which the amount of recombination represents a minimum estimate, since some exceptional animals die before sufficient offspring can be ob-

tained to classify them definitely as recombinants or overlaps. It is nevertheless evident that *T*, *Ki* and *Fu* are closely linked, probably all within less than three crossover units. The mutations at these three loci are thus not only phenotypically similar but seem to be near neighbors in the chromosome. We are led, therefore, to inquire whether their similarities and their proximity in the chromosome may be related.

It is conceivable, of course, that the two circumstances have no causal relation at all; and that by chance the mutations occurring at three adjacent loci happen to have similar effects. The likelihood of such a coincidence can be estimated by assuming the ten loci in the mouse known to have produced mutations of a similar type§ to be distributed at random among the 20 chromosomes. The chance that any chromosome should have one of these loci is thus  $10/20$  or  $1/2$ ; the chance that one chromosome should have 3 such loci is  $(1/2)^3$ . The genetic lengths, in crossing-over units, of the chromosomes of the mouse have not been determined. However, the average number of chiasmata per chromosome may be estimated as just over two (Crew and Koller, 1932) and this would lead to an estimated average map length of about 100. Assuming, therefore, that a chromosome consists of 25 crossover segments of about four crossover units each, and that crossing-over in this chromosome is normal, the chance that three loci chosen at random should fall within one segment would be  $(1/25)^3$ . Consequently, the likelihood that three loci should have the distribution actually found in our sample would be  $(1/2)^3 \cdot (1/25)^3$  or 0.000008. This is rather remote; and it would seem more likely that the similar effects and near location are related to each other.

The simplest assumption concerning this relationship would be that *T*, *Fu* and *Ki* represent mutations in duplicate loci lying next to each other in the chromosome in the manner of the "repeats" discovered in *Drosophila* by Bridges (1935). Lewis (1942) has recently shown that Star and asteroid, two phenotypically similar and very closely linked mutations in *Drosophila melanogaster* are located in one double band of the salivary chromosome, the doubling having presumably occurred through repetition of one band. Although such a repeat hypothesis would reconcile the phenotypic similarity and the close linkage existing among the three dominants, it fails to account for the recessives at the *T* locus ( $t^0$ ,  $t^1$ ) and the fact that  $t^0 t^0$ ,  $t^1 t^1$ , *T T* and *Ki Ki* are lethal; that is, these recessive effects are not "covered" as they should be if normal duplicate loci are present. Although the duplication hypothesis might be rescued by assuming that the duplicated normal loci have lost, by virtue of their position, that portion of their normal effect which "covers" the lethal effect, we have no independent evidence that this is so. It is possible, however, to test one consequence of the duplication hypothesis, namely, that mutations at repeated or duplicated loci should retain some basic similarity in developmental effect like

that which marks the mutations at the same locus. To this end, the effects of *Fu* and *Ki* on embryological processes are being compared with those already known to be associated with *T*,  $t^0$  and  $t^1$  (see Dunn, 1941, for review).

It is, of course, possible to conceive of a developmental unit in the chromosome having greater extension than one crossing-over unit and to regard the manifestations of these several mutations as effects of such a larger unit. However, until it is known whether the developmental effects are unitary or diverse such ideas will lack support. It will also be necessary to test whether crossing-over is normal in the chromosome involved, since it is possible that the small amount of recombination between loci *T*, *Fu* and *Ki* is due to some more general suppression of crossing-over. If the developmental processes chiefly affected turn out to be not as similar as the adult phenotypes seem to indicate or if crossing-over is not normal, then chance association of three similar mutations would again have to be considered as a possible explanation.

\* It is a pleasure to acknowledge the help of Dr. Salome Gluecksohn-Schoenheimer and the financial assistance of the fund for research of Columbia University.

† Aided by a grant of the John and Mary R. Markle Foundation.

‡ Since this shows Brachy and Fused to be at separate loci, we shall adopt the symbol *Fu* replacing  $T^F$  for the Fused locus, retaining *T* for Brachy.

§ The ten loci are: *T*,  $t^0$ ,  $t^1$  (Dunn, 1941); *Fu* (this paper); *Ki* (Caspari and David, 1940); *Sd* (Dunn, Gluecksohn-Schoenheimer and Bryson, 1940); *st* (Dunn, 1934); *sb* (stub) (Dunn and Gluecksohn-Schoenheimer (in press)); *tw* (twist) (Dunn and Gluecksohn-Schoenheimer, unpublished); *sc* (screw) (Laanes and MacDowell, 1942); *fl* (flex) (Hunt, *et al.*, 1933); *pt* (pigtail) (Crew, 1941).

Bridges, C. B., *Jour. Heredity*, **26**, 60-64 (1935).

Caspari, E., and David, P. R., *Ibid.*, **31**, 427-431 (1940).

Chesley, P., *Jour. Exp. Zool.*, **70**, 429-459 (1935).

Chesley, P., and Dunn, L. C., *Genetics*, **21**, 525-536 (1936).

Crew, F. A. E., and Auerbach, C., *Jour. Genetics*, **41**, 267-274 (1941).

Crew, F. A. E., and Koller, P. Ch., *Ibid.*, **26**, 359-383 (1932).

Dobrovolskaia-Zavadskaja, Kobozeff, N., and Veretennikoff, S., *Archives de Zoologie Experimentale*, 1934, pp. 250-358.

Dunn, L. C., these PROCEEDINGS, **20**, 230-232 (1934).

Dunn, L. C., *Genetics*, **24**, 728-731 (1939).

Dunn, L. C., *Third Growth Symposium*, 1941, pp. 147-161.

Dunn, L. C., and Gluecksohn-Schoenheimer, S., *Genetics*, **23**, 146-147 (1938).

Dunn, L. C., Gluecksohn-Schoenheimer, S., and Bryson, V., *Jour. Heredity*, **31**, 343-348 (1940).

Hunt, H. R., Mixter, R., and Permar, D., *Genetics*, **18**, 335-366 (1933).

Laanes, T., and MacDowell, E. C., *Ibid.*, **27**, 151-152 (1942).

Lewis, E. B., *Ibid.*, **27**, 153-154 (1942).

Reed, S. C., *Ibid.*, **22**, 1-13 (1937).



## THE EFFECT OF pH ON INACTIVATION OF TOBACCO MOSAIC VIRUS BY X-RAYS\*

BY A. MARSHAK† AND WILLIAM N. TAKAHASHI

RADIATION LABORATORY AND THE DIVISION OF PLANT PATHOLOGY,  
UNIVERSITY OF CALIFORNIA

Communicated April 13, 1942

Previous experiments have shown that the sensitivity of chromosomes to x-rays can be reduced by treatment with ammonium hydroxide and that the protective action of the ammonium hydroxide increases with concentration.<sup>6</sup> This effect was attributed to the removal of positive charges on the chromosome surfaces. The experiments with tobacco mosaic virus reported here show that on the acid side of the isoelectric point where the molecules carry a net positive charge they are more sensitive to x-rays.

*Experimental.*—The purified virus, very kindly given to us by Dr. W. M. Stanley, was isolated by ultracentrifugation. One-tenth of a cc. of a distilled water solution containing 2.1 mg. of the virus was suspended in nine-tenths of MacIlvaine buffer at the proper pH and irradiated in paraffin-lined celluloid capsules. One sample of virus in buffer at pH 7.0 (reference point) was always irradiated at the same time as samples at other pH values. After irradiation the volume of solution was made up to 100 cc. and adjusted to pH 7.0 with the appropriate buffer solution. Controls received the same treatment except that irradiation was omitted. Experiments were limited to pH values from 2.2–7.0 since investigations of other workers<sup>9</sup> have shown that the virus is inactivated beyond that range.

The relative concentration of active virus was determined by the usual biological method of inoculating opposite halves of *Nicotiana glutinosa* leaves<sup>4,7</sup> with the irradiated virus and unirradiated controls. For each sample 20 half-leaves were used.

In a preliminary experiment portions of a single sample of virus in distilled water were given different doses of x-rays and tested on half-leaves after diluting 1/1000. The curve of the logarithm of the number of lesions produced by irradiated virus as plotted against per cent of the unirradiated control is given in figure 1. The results fit the equation  $Y = e^{-kx}$ , where  $Y$  is the per cent unaltered virus,  $k$  a constant and  $x$  the dose in Roentgens.  $k$  has the value of  $8.12 \times 10^{-6}$  and the six points obtained fit the curve to within 1%. The sample of virus irradiated two weeks later showed inactivation within 2% of the values previously obtained. However, another sample of virus prepared in the same manner showed much greater sensitivity to x-rays. Curve *a* of figure 1 gives the results obtained with the first sample and curve *b* those of the second. The slopes of the two curves and therefore the x-ray sensitivity of the two samples of virus



differ by a factor of approximately 9. It was found necessary to use different samples of virus in the course of the first pH experiment. This, as will be seen later, did not influence the results obtained.

*Results.*—The counts of local lesions on pairs of half-leaves are listed in table 1.

TABLE 1

*N. glutinosa* LESION COUNTS FOLLOWING IRRADIATION AT DIFFERENT pH VALUES

pH	$C_7$	(1) $I_7$	$C_x$	(2) $I_x$	$C_7$	(3) $C_x$	$I_7$	(4) $I_x$
2.2	2015	1334	2025	1023	2430	2265	1002	613
3.0	2066	1344	2255	1387	2602	2457	1097	922
4.0	1213	867	1397	1017	(1)1154 (2) 433	1010 442	744	702
5.0	1854	1361	1537	1048	1486	1312	1308	1229
6.0	1854	1361	1707	1344	1611	1456	1445	1442

Column 1 gives the number of lesions produced by unirradiated virus at pH 7.0 ( $C_7$ ) and irradiated virus at pH 7.0 ( $I_7$ ), each being on halves of the same leaf of *N. glutinosa*. Similarly the spots produced on one set of half-leaves by unirradiated virus at different pH values is given by  $C_x$  (column 2) and the spots produced on corresponding half-leaves by virus irradiated at different pH concentrations by  $I_x$ . The controls in column 3 are compared with the irradiated samples in column 4.

The fraction of non-inactivated virus at pH 7.0 for any one run is given by  $I_7/C_7$  and similarly at other pH concentrations by  $I_x/C_x$ . Letting  $I_7/C_7 = S_7$  and  $I_x/C_x = S_x$ , then  $S_x/S_7$  will give the survivors at pH<sub>x</sub> as a fraction of the survivors at pH<sub>7</sub>, and  $1 - S_x/S_7$  will give the efficiency of the irradiation in inactivating the virus at pH<sub>x</sub>. The results of these computations are given in column a table 2. The efficiency of the x-rays at different pH concentrations may also be calculated from the data of (3) and (4) of table 1. The virus inactivation by pH alone will be given by  $1 - (C_x/C_7)$  and this subtracted from  $1 - (I_x/I_7)$  will give the x-ray efficiency of pH<sub>x</sub>.

TABLE 2

THE RELATIVE EFFICIENCY OF X-RAYS AT DIFFERENT pH VALUES

pH	a $\left(1 - \frac{S_x}{S_7}\right) 100$	b $\left[\left(1 - \frac{I_x}{I_7}\right) - \left(1 - \frac{C_x}{C_7}\right)\right] 100$
2.2	+23	+32
3.0	+10	+10
4.0	- 2	- 7, +6
5.0	+ 7	- 6

Table 2 gives a summary of the x-ray efficiency in per cent of the inactivation at pH 7.0. At pH 4.0 a second set of inoculations was made with  $C_x$  and  $C_7$  with the result shown in table 1. It is clear from this and

from the differences in the values obtained by the two methods described that errors as large as 13% may be expected.

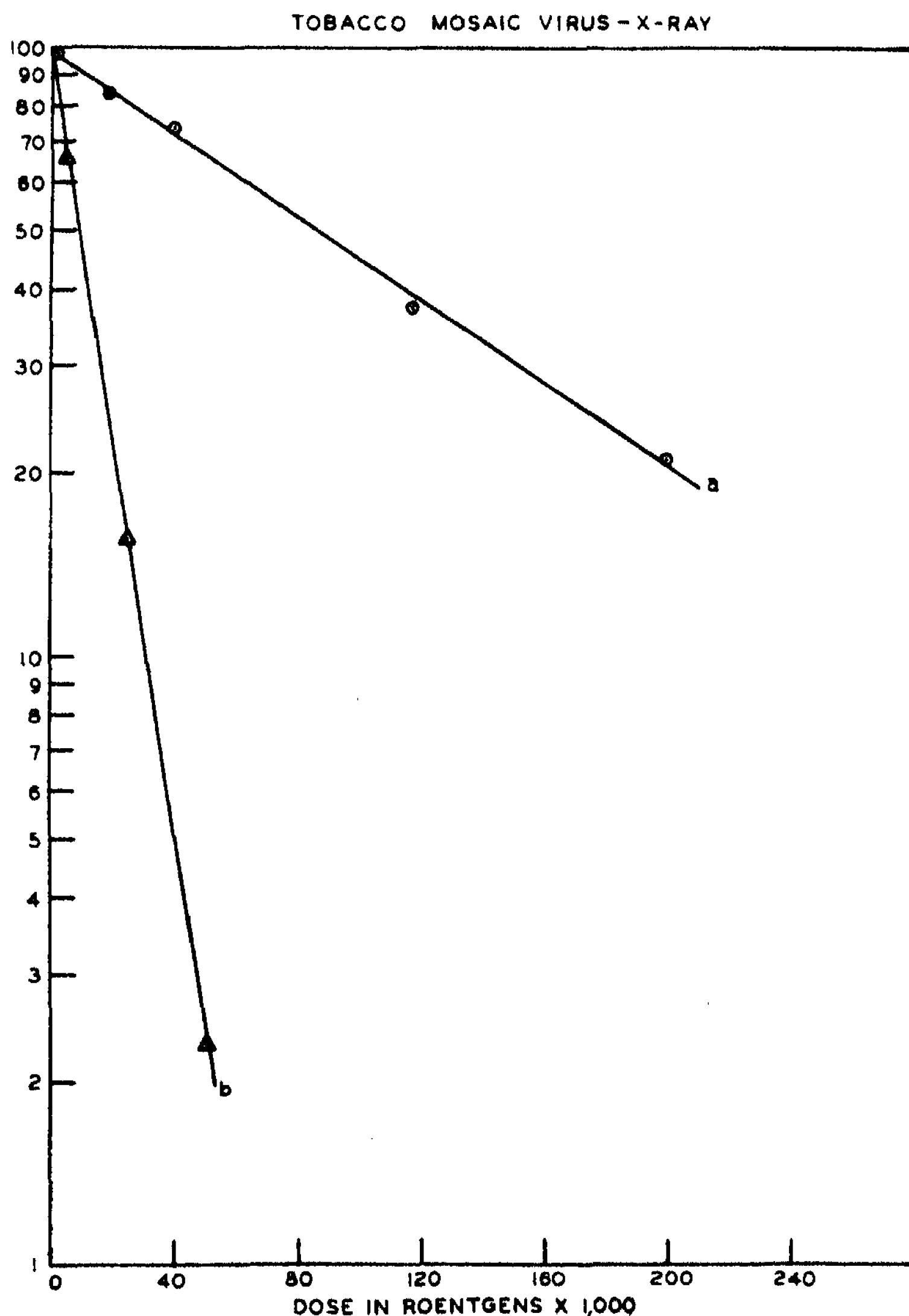


FIGURE 1

From examination of table 2 it appears that the efficiency of x-rays between pH 4.0 and 6.0 is equal to or less than the efficiency at pH 7.0. The differences are not large enough to be significant. However, at pH 2.2-3.0 the efficiency is significantly increased. To determine whether this might

have been an accidental result two more experiments were performed. In both of these the same sample of virus was used in determining inactivation at all pH values used. The results shown in table 3 are essentially similar to those of the previous experiment. At pH 2.2 x-rays are 25–30% more efficient in inactivating the virus than at pH 7.0, and at pH 3.4 the x-ray efficiency is greater by 10–15%. Between pH 3.4 and pH 7.0 there are no significant differences in the results.

TABLE 3

*N. glutinosa* LESION COUNTS AND X-RAY EFFICIENCY AT DIFFERENT pH VALUES

		$C_{2.2}$	$I_{2.2}$	$C_{3.4}$	$I_{3.4}$	$C_{4.6}$	$I_{4.6}$	$C_{7.0}$	$I_{7.0}$
Number	1	608	276	515	280	298	207	425	276
	2	220	110	266	156	322	207	284	199
$I$	1	0.454		0.544		0.695		0.650	
$\bar{C}$	2	0.50		0.586		0.643		0.678	
$(1 - \frac{S_2}{S_1})$	100								
	1	+30		+16		-7		0	
	2	+26		+13		+5		0	

Gowen<sup>3</sup> studied the effect of x-rays on tobacco mosaic virus and found exponential survival curves. From these curves he calculated the size of the virus to be  $7.5 \times 10^{-18}$  cm.<sup>3</sup>. He does not report differences in survival curves with different samples of virus. The sensitive volumes calculated from curves *a* and *b* of our experiments are  $4.6 \times 10^{-18}$  cm.<sup>3</sup> and  $4.2 \times 10^{-17}$  cm.<sup>3</sup>, respectively. Lea and Smith<sup>5</sup> irradiated dried and aqueous suspensions of tobacco necrosis virus with x-rays and found the results of both methods to fall on the same exponential survival curves. They mention that in different experiments different rates of inactivation were observed and the comparison mentioned above is made only with what they consider their best results. They found that tobacco mosaic virus in different states of aggregation gave essentially similar survival curves when treated with ultra-violet light. However, the experimental error of their observations was sufficiently large to make it impossible to determine whether their data indicated a response by elementary particles or aggregates as high as 4.

In the present experiments the shape of the survival curve is determined with an error no greater than 1–2% and can be explained only in terms of a response to x-rays by a single functional virus unit. Pirie<sup>1</sup> and Frampton<sup>2</sup> have presented evidence indicating that the virus particles may exist as aggregates. Filtration studies of the tobacco necrosis virus by Smith and MacClement<sup>8</sup> indicate the presence of aggregates in extracts containing this virus. The form of the survival curve we have obtained requires that at least one functional unit of the virus be inactivated for every ion pair or cluster produced in the virus elementary particle or aggregate. If aggregates exist, the following possible interpretations may be given to the data:

(1) There is only one functional unit in the aggregate; the rest is inert material.

(2) Aggregates are not permanent but are continually being broken down and built up from elementary units. As pointed out by Lea and Smith in this case the inactivation of particles in aggregates will proceed at the same rate as the inactivation of elementary particles and simple exponential curves will be obtained.

In both (1) and (2) we may assume that:

(a) The energy released by an ion pair or cluster within one elementary unit will inactivate only that unit.

(b) The energy released by an ion pair or cluster will inactivate more than one unit.

If after irradiation the virus is diluted before inoculation the number of aggregates will be decreased. Assuming condition (1), the apparent sensitivity of the virus will be independent of the state of aggregation and the spread of energy through the aggregate could not be detected. Under condition (2a) the sensitivity of the virus will be independent of the state of aggregation, while under condition (2b) the greater the aggregation at the time of the irradiation the greater the apparent sensitivity. The results obtained point to the latter hypothesis.

If it is postulated that the greater sensitivity to x-rays at pH 2.2 to 3.4 is due to a greater number of aggregates than at pH 3.4–7.0 the results cannot be adequately explained. One would expect a maximum sensitivity at pH 3.4, the isoelectric point, whereas the maximum observed is at 2.2. One would also expect that the samples of virus containing different amounts of aggregates would have different relative responses to x-rays at the various pH values. However, the same response was obtained although the sensitivity of the samples varied by a factor of 9. From these considerations it follows that the pH effect on x-ray sensitivity of the virus cannot be explained in terms of differing states of aggregation of the virus. A consideration of the charges on the suspended particles does provide an adequate explanation. The more acid the suspending medium, the greater will be the net positive charge on the suspended particles. The positively charged member of an ion pair produced by x-rays being of atomic size will be prevented from reaching the virus particle before losing its charge, while the negative ion being only of electronic mass will be able to penetrate to the virus and carry sufficient energy to inactivate the virus particle. The greater the net positive charge on the virus the greater will be its attraction for the free electrons produced by x-rays.

*Conclusion.*—1. Energy released by an ion pair produced within a virus particle may travel through that particle and inactivate one or more of the elementary virus units it may contain.

2. Electrons reaching the virus from the suspending medium have sufficient energy to inactivate the virus particle.

3. More of the electrons produced by x-rays in the suspending medium reach the virus particles when the latter carry a net positive charge than when the charge is negative. The greater the net positive charge the more electrons will be attracted.

\* This work has been supported by the Columbia Foundation.

† Fellow of the Finney-Howell Cancer Research Foundation.

<sup>1</sup> Bawden, F. C., and Pirie, N. W., *Proc. Roy. Soc. (B)*, **123**, 274 (1937).

<sup>2</sup> Frampton, V. L., *Science*, **95**, 232 (1942).

<sup>3</sup> Gowen, J. W., *Proc. Nat. Acad. Sci.*, **26**, 8 (1940).

<sup>4</sup> Holmes, Francis O., "Local Lesions in Tobacco Mosaic," *Bot. Gaz.*, **87**, 39 (1929).

<sup>5</sup> Lea, D. E., and Smith, K. M., *Parasitology*, **32**, 405 (1940).

<sup>6</sup> Marshak, A., *Jour. Gen. Physiol.*, **19**, 179 (1935); *Proc. Soc. Exp. Bio. & Med.*, **38**, 705 (1938); *Ibid.*, **39**, 194 (1938).

<sup>7</sup> Samuel, Geoffrey, and Bald, J. G., "On the Use of Primary Lesions in Quantitative Work with Two Plant Viruses," *Ann. Appl. Biol.*, **20**, 70 (1933).

<sup>8</sup> Smith, K. M., and MacClement, W. D., *Parasitology*, **32**, 320 (1940).

<sup>9</sup> Wyckoff, R. W. G., *Jour. Biol. Chem.*, **122**, 239 (1937).

PROCEEDINGS  
OF THE  
NATIONAL ACADEMY OF SCIENCES

Volume 28

June 15, 1942

Number 6

Copyright 1942 by the National Academy of Sciences

*THE RELEASE OF AUXIN FROM ISOLATED LEAF PROTEINS  
OF SPINACH BY ENZYMES*

By SAM G. WILDMAN AND SOLON A. GORDON\*

DEPARTMENT OF BOTANY, UNIVERSITY OF MICHIGAN†

Communicated March 25, 1942

A number of hypotheses have been advanced concerning the substance from which auxin is liberated in plant tissues. Kögl, Erxleben and Haagen Smit<sup>9</sup> postulated the existence of auxin esters in fats and oils, since they were able to hydrolyze auxin from oil of *Arachis* and other plants by means of lipase or saponification with alcoholic sodium hydroxide. Recently there appeared a brief note by Skoog and Thimann<sup>13</sup> to the effect that an increase of auxin could be obtained from *Lemna* tissue by the addition of proteolytic enzymes. They considered that the auxin was bound to proteins. Avery, Berger and Shalucha<sup>1</sup> have criticized this hypothesis as being premature since proteolytic enzymes also hydrolyze simple peptides and amides. In addition, they were unable to hydrolyze the auxin precursor in a corn endosperm extract with commercial preparations of trypsin, pepsin, papain, ficin, taka-diastase, malt diastase or steapsin at pH values of 7 or less although the precursor could be hydrolyzed by other means. They suggest that the precursor in maize may be an acid amide of indole acetic acid. It is at once evident that the information concerning this phase of plant hormone research is extremely meager.

Methods for the isolation of proteins from leaves have been proposed by Chibnall,<sup>3</sup> Menke<sup>11</sup> and Granick,<sup>5, 6</sup> and it should be possible to test the hypothesis that auxin is associated with protein by the use of isolated plant proteins. If it can be shown that auxin is definitely a part of protein molecules or associated with them, some ramifications of the problem may be studied such as (1) whether auxin is confined to cytoplasmic or chloroplastic protein, or found in both; (2) if it is found in leaf proteins, can these proteins be further fractionated so that the yield of auxin from one fraction is greater than from another; (3) diffusion rates of the auxin released from proteins can be made to determine whether they compare with the values of auxin directly extracted from the leaf. With these considerations in mind, the following experiments were performed.

*Preparation of Proteins.*—*Cytoplasmic proteins:* Chibnall<sup>8</sup> has shown that vacuolar materials, and the soluble substances of the cytoplasm could be removed by cytolysis of the leaf tissue followed by pressure to "squeeze" the cell fluids from the leaf, cytoplasmic protein and chloroplasts remaining within the cell walls. The cells were then broken by grinding, the ground material suspended in water and the cell wall fragments removed by passing the liquid through silk. The chloroplasts were separated from the cytoplasmic proteins by passing the suspension through paper pulp. Menke,<sup>11</sup> on the contrary, ground spinach leaves without previous treatment and separated the chloroplastic protein from the cytoplasmic protein by precipitation with 25 per cent ammonium sulphate.

A combination and modification of the methods proposed by Chibnall and Menke have been used to prepare the leaf proteins of spinach for this study. Approximately one kilogram of market spinach (consisting only of leaves) was thoroughly washed and cytolized for a few minutes with ether. The flaccid leaf material was surrounded by cotton cheesecloth and the juice squeezed out in a hydraulic press. Upon releasing the pressure, the leaves were allowed to imbibe water and then pressed again. Imbibition and pressing were repeated three times. After the final pressure treatment, the material was again allowed to imbibe water and the leaves were passed twice through the coarse burr of a "Nixtamal" mill, then three times more using the finest burr. Following this preliminary grinding, water was added in small quantities and the leaf material was repeatedly passed through the mill until a semi-liquid consistency was obtained. The finely divided material was then dispersed in about two liters of distilled water. Cell wall debris was removed by a short centrifugation at low speed and the supernatant liquid was passed through fine mesh silk to remove all visible traces of cell fragments. The combined centrifugates were again ground and dispersed as before and both dispersions combined to give a total of about three liters of liquid.

The chloroplasts were precipitated by adding ammonium sulphate to give a final concentration of 7.5 per cent. The material was allowed to flocculate for 12 hours at 5°C., and the precipitated chloroplasts removed by filtering through Whatman No. 12 folded filter paper. The filtrate was returned through the same paper until the filtrate became free from green coloration. All filtering was done at 5°C. with toluol present to prevent bacterial decomposition of the proteins. We have used 25, 20, 15, 10 and 5 per cent ammonium sulphate to precipitate the chloroplasts in other preparations. Table 1 expresses the nitrogen (and protein) content of some of these preparations. Nitrogen was determined by the micro-Kjeldahl method using selenium as a catalyst. However, smaller yields of cytoplasmic protein than anticipated seemed to indicate that some cytoplasmic protein precipitated with the chloroplasts when high concentrations of

ammonium sulphate were used; hence, the values reported here are confined to those preparations in which not more than 15 per cent ammonium sulphate was used.

TABLE 1

NITROGEN CONTENT OF PROTEINS USED FOR AUXIN ANALYSIS. PROTEINS WERE SUBJECTED TO PROLONGED EXTRACTION WITH ETHER AND ETHER-ALCOHOL BEFORE ANALYSIS (SEE TEXT)

KIND OF PROTEIN	% (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> USED TO PRECIPITATE CHLORO- PLASTIC PROTEINS	% N	% PROTEIN (N × 6.25)
Cytoplasmic, precipitated at pH 4.0	15.0	15.6	97.5
	15.0	15.5	96.9
Cytoplasmic, precipitated at pH 4.0	10.0	15.7	98.1
	10.0	15.6	97.5
Cytoplasmic, precipitated at pH 5.2	7.5	15.4	96.3
	7.5	15.5	96.9
Cytoplasmic, precipitated at pH 3.5	7.5	15.5	96.9
Cytoplasmic, precipitated at pH 5.2	5.0	15.2	95.5
	5.0	15.2	95.5
Chloroplastic, prepared by Granick's method	..	11.1	69.4
	..	10.9	68.3

Cytoplasmic protein contained in the filtrate was precipitated by adjusting the pH of the solution to 4.0 with 0.1 *N* hydrochloric acid. All pH measurements were made with a Beckmann glass electrode. The precipitate was allowed to flocculate for several hours and then collected on Whatman No. 1 filter paper. The protein precipitate was washed until free from sulphate with distilled water whose pH had been adjusted to 4.0. The precipitate was then removed from the filter paper and dried in vacuum at 37°C. The dried proteins were ground to pass through a 100-mesh screen and extracted with ether for 72 hours in a Soxhlet apparatus. This extraction was followed by a 1:4 ether-alcohol extraction for 72 hours. Both the ether and ether-alcohol extracts gave negative tests for auxin by the *Avena* assay method. A yield of four grams of cytoplasmic protein has been obtained from one kilogram of leaf tissue when the chloroplasts were precipitated with 7.5 per cent ammonium sulphate.

*Chloroplastic proteins:* With the object of obtaining a relatively pure sample of chloroplasts, a modification of the differential-centrifugation method suggested by Granick<sup>5, 6</sup> was used. Seventy-five grams of market spinach were macerated to yield a suspension in glucose solution. Centrifuging seven minutes at 700 RCF threw out the heavy cell debris, crystals, starch granules and clumps of chloroplasts. The supernatant liquid was decanted and centrifuged ten minutes at 1150 RCF to yield a centrifugate with a white layer, probably suspended sand and porcelain, at its lower portion and covered by a layer of chloroplasts. The supernatant liquid of this



deposit was centrifuged fifteen minutes at 1490 RCF to yield a centrifugate of dark green color. Microscopic examination of this showed mostly isolated chloroplasts with an occasional fragment. To remove the glucose, the chloroplasts were suspended and centrifuged twice in distilled water fifteen minutes at 1490 RCF. Then, they were dried *in vacuo* at 37°C., ground to 100 mesh and extracted successively with ether and ether-alcohol. The nitrogen content of this preparation is given in table 1.

*Digestions of Proteins by Enzymes.*—All digestions of protein herein reported have been carried out under constant conditions. Preliminary experiments indicated that the maximum rate of digestion with tryptic enzymes occurred between a pH of 7.0 and 9.0, using a  $\text{KH}_2\text{PO}_4$ -NaOH buffer medium. It was likewise found that 72 hours was the minimum time that could be used to digest the proteins and still obtain consistent yields of auxin. Inasmuch as proteolytic enzymes of animal origin appear to attain their maximum efficiency at temperatures near that of blood,<sup>4</sup> an incubation temperature of 37°C. was used. Digestions were made with 10 mg. of protein suspended in 10 ml. of buffer solution at pH 8.0 with the exception of those digestions employing papain in which 10 ml. of phosphate buffer at pH 4.5 and 6.0 were used. Four to five drops of toluol were added to each digestion flask as an antiseptic and renewed during the digestion period if necessary; each flask was tightly stoppered to prevent evaporation of the toluol. At the completion of the 72-hour digestion period, the pH of the solution was adjusted to 4.0–4.5, the solution transferred to a separatory funnel and shaken twice with 75 to 100-ml. portions of recently distilled ether. Ether containing the auxin was separated from the protein solution and evaporated at low temperature until only a few milliliters remained. These were transferred to a small vial to which agar was added.

*Auxin Determinations.*—Auxin was determined by the standard *Avena* test procedure,<sup>16</sup> using a reaction time of 90 minutes and a uniform intermediate decapitation time of 2–2.5 hours. All values herein reported are based on determinations made within the dilution range. Dilutions were made either by diluting the agar which contained the hormone, or by dissolving the extract of the hormone in ether, making up to volume with ether at 5°C., and removing suitable aliquots. Although van Overbeek's<sup>12</sup> formula is of utility in *Avena* determinations, the constants involved when 10 to 40-mg. samples are used are of such magnitudes as to make its use impractical in expressing the results obtained here. Hence, the values expressed are actual *Avena* curvatures per 0.35 ml. of agar and comparisons are based upon determinations made on the same day to obviate correcting for differences in sensitivity of the test plants.

*Enzymes Used.*—We have repeatedly tested various concentrations of the enzymes used for the presence of auxin; the *Avena* assay was always nega-

tive. They were tested before and after 72 hours' incubation under conditions identical with the conditions of protein digestions.

Trypsin and chymotrypsin were commercially prepared crystalline enzymes. Papain† was a twice crystallized preparation dissolved in 0.2M NaCN and had an activity of 10 milk-clotting units per 3.7 mg. of protein. What is called tryptic extract in this paper is a highly purified, non-crystalline trypsin commercially prepared from pancreas and presumably containing a mixture of proteolytic enzymes.<sup>14</sup> Preliminary experiments disclosed no significant difference in auxin yield when 0.5, 1.0 and 5.0 mg. of crystalline enzymes per 10 mg. of protein were used; therefore, 1.0 mg. of crystalline enzyme was used for each digestion with trypsin or chymotrypsin. Similarly, 1.0, 5.0 and 10.0 mg. of tryptic extract were not markedly different in ability to yield auxin. Accordingly, 5.0 mg. of enzyme was added to each digestion in which tryptic extract was used. 1.0 mg. of papain was used in those digestions entailing this enzyme.

*Bacterial Contamination.*—The possibility exists that auxin liberated from protein may be due to the action of bacteria, either on the protein or on the hydrolytic products of the protein. However, we feel that the following considerations will substantiate our later thesis that the auxin is released by enzymatic hydrolysis and not through the functioning of microorganisms: (1) Toluol was employed as an antiseptic. Any digestion in which the odor of toluol could not be clearly detected at the time of neutralization was discarded. (2) Microscopic examination of the clear liquid from digestions to which toluol had been added disclosed no microorganisms; when toluol was omitted, the liquid became cloudy and its odor offensive; microscopic inspection revealed numerous organisms. (3) No turbidity or odor had developed in a tightly stoppered flask containing 40 mg. of protein, 40 ml. of buffer, 4.0 mg. of chymotrypsin and four drops of toluol at the end of three months. (4) The yields of auxin obtained from like protein and enzyme in various experiments were markedly uniform. That this reproducibility should exist with fortuitous contamination of the digestions by air-borne spores is not probable. (5) Twelve protein digestions were prepared as previously described with trypsin, chymotrypsin and tryptic extract in the presence of toluol. After 72 hours' incubation, 1 ml. of each was plated out upon beef-peptone agar using sterile transfer technique. A few drops of toluol were then added to eight of the plates, none being added to the remaining four. No bacterial or fungus colonies developed within ten days in those plates protected with toluol. Conversely, numerous colonies developed within four days on the four plates not protected by toluol. Undoubtedly the digestions do contain a few microorganisms, but we are convinced that toluol keeps them inactive.

*Experimental Results: Hydrolysis of Proteins to Yield Auxin.*—Examination of the data presented in table 2 clearly demonstrates that auxin is

associated with protein, and is released through the action of proteolytic enzymes. Like results have also been obtained for three other protein preparations, and always there has been complete corroboration of the data presented. Since the dry protein was suspended in buffer solution, liquefaction of the protein and attainment of a clear solution is assumed to represent hydrolysis. Protein controls, to which no enzyme had been added, failed to dissolve completely during a period of 72 hours, although there sometimes appeared to be a small amount of hydrolysis as attested by *Avena* curvatures. Chibnall has shown that there is a gradual auto-digestion when cytoplasmic proteins are maintained in aqueous solution or in a highly hydrated form. The inference is that intracellular enzymes are present and accomplish this hydrolysis. Thus, one would expect a small amount of auxin to be normally produced in the absence of added enzyme.

TABLE 2

AUXIN RELEASED FROM TWO DIFFERENT PREPARATIONS OF CYTOPLASMIC PROTEINS. DUPLICATES ARE THE RESULT OF TWO SEPARATE *Avena*-CURVATURE DETERMINATIONS WITH TWO SEPARATE PROTEIN DIGESTIONS

PROTEIN PREPARATION	NO. OF 10-MG. DIGESTIONS	ENZYME	AVERAGE <i>Avena</i> CURVATURE IN DEGREES
Cytoplasmic proteins precipitated at pH 4.0, after precipitation of chloroplastic proteins by 10% (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	4*	Control	0
			2.0 ± 0.3
	4	Chymotrypsin	4.4 ± 0.4
			3.8 ± 0.5
	4	Trypsin	6.0 ± 0.4
			4.8 ± 0.6
	4	Tryptic extract	15.0 ± 0.6
			12.5 ± 0.7
Cytoplasmic proteins precipitated at pH 4.0, after precipitation of chloroplastic proteins by 15% (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	4	Control	0
			0
	4	Chymotrypsin	2.0 ± 0.9
	4	Trypsin	0
	4	Tryptic extract	8.4 ± 0.8
Cytoplasmic proteins precipitated at pH 5.2, after precipitation of chloroplastic proteins by 7.5% (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>			9.1 ± 0.4
	3	Control	0
	4	Control	0
			1.3 ± 0.5
	4	Trypsin	0
	2	Chymotrypsin	0
	4	Chymotrypsin	0
	1	Tryptic extract	1.0 ± 0.3
			2.6 ± 0.8
	3	Tryptic extract	2.9 ± 0.6
	4	Tryptic extract	8.5 ± 1.3

\* Ether extracts of four separate digestions combined.

When the three enzymes are compared as to their efficacy in causing auxin to be released from protein, tryptic extract (so far as our data go) is the best. Trypsin and chymotrypsin are less effective. Little, if any, visible hydrolysis of proteins resulted when papain was used either at pH 4.0 or 6.0. The results were inconclusive. For instance, one experiment would show an apparently significant increase in auxin but an attempt to duplicate it would fail. The amount of papain available was insufficient to determine the optimum conditions for the hydrolysis of leaf proteins. The most significant fact here, however, is the reproducibility of the results: two separate digestions of equal weight of protein yield nearly equal amounts of auxin as measured by *Avena* curvatures.

*Isoelectric Fractionation of Cytoplasmic Proteins.*—After the chloroplasts had been removed by precipitation with ammonium sulphate and filtering, 0.1 *N* hydrochloric acid was added to a filtrate containing cytoplasmic proteins until a pH of 5.2 was attained. At this hydrogen-ion concentration, a rapid and voluminous flocculation of protein occurred. Flocculation

TABLE 3

AUXIN YIELD FROM TWO FRACTIONS OF CYTOPLASMIC PROTEINS OBTAINED BY ISO-ELECTRIC PRECIPITATION. (CHLOROPLASTIC PROTEIN HAD BEEN SEPARATED FROM CYTOPLASMIC PROTEIN BY 7.5%  $(\text{NH}_4)_2\text{SO}_4$ )

NO. OF 10-MG. DIGESTIONS	ENZYME	AVERAGE <i>Avena</i> CURVATURE IN DEGREES PROTEIN PREPARATION	
		PRECIPITATED AT pH 5.2	PRECIPITATED AT pH 3.5
4	Control	1.3 $\pm$ 0.5	5.2 $\pm$ 0.7
1	Control	0	0
1	Chymotrypsin	0	9.8 $\pm$ 1.2
2	Chymotrypsin	0	16.8 $\pm$ 1.0
1	Trypsin	0	6.2 $\pm$ 0.7
		0	1.4 $\pm$ 0.2
1	Tryptic extract	1.0 $\pm$ 0.3	15.3 $\pm$ 1.1
		1.3 $\pm$ 0.4	

was allowed to proceed at room temperature for one-half hour, and the precipitate was then removed by filtration at 5°C. which required about six hours. The filtrate was then adjusted to pH 3.5 which precipitated more protein. This was removed by filtration. Both precipitates were thoroughly washed with water properly acidified, dried *in vacuo* at 37°C. and extracted with ether and ether-alcohol. The protein fraction precipitated at pH 3.5 constituted about five per cent of the total cytoplasmic protein. Auxin yields obtained by enzymatic hydrolysis of these two cytoplasmic protein fractions appear in table 3.

When considered on a basis of protein weight, it will be noticed that the protein fraction precipitated at pH 3.5 contains considerably more auxin than the protein precipitated at pH 5.2. Here again, the same trend is evident: tryptic extract releases more auxin than trypsin or chymotrypsin.

The presence of intracellular enzymes capable of hydrolyzing proteins and releasing auxin is suggested by the relatively large curvature obtained in the control without enzymes using four 10 mg. digestions of protein. However, the control consisting of one 10 mg. digestion of protein should be used for comparison with those digestions to which enzymes were added.

*Auxin Yield from Chloroplastic Proteins.*—The values in table 4 disclose that auxin is also obtained by enzymatic hydrolysis from a protein preparation of spinach chloroplasts. From table 1, it is obvious that the chloroplastic protein preparation is of lower purity than the cytoplasmic protein preparations. If the curvatures in table 4 are weighted on a per-milligram-of-nitrogen basis, a better comparison can be obtained between the auxin content of the two proteins. Of course, this is made on the assumption that auxin is associated only with the protein of the chloroplasts and cytoplasm. The values for auxin obtained from chloroplastic protein were increased by 40 per cent; table 5 shows these curvatures weighted to allow for impurities. Compensating for the impurity in the chloroplastic protein, the two proteins yielded approximately the same amount of auxin with trypsin and tryptic extract, but not with chymotrypsin.

TABLE 4  
AUXIN YIELD FROM CHLOROPLASTIC AND CYTOPLASMIC PROTEINS

NO. OF 10-MG. DIGESTIONS	ENZYME	AVERAGE <i>Avena</i> CURVATURE IN DEGREES	
		CHLOROPLASTIC PROTEIN (BY GRANICK'S METHOD)	CYTOPLASMIC PROTEIN PRECIPITATED AT pH 8.5
2	Control	2.6 $\pm$ 0.6	3.5 $\pm$ 0.4
1	Control	0	0
1	Trypsin	3.9 $\pm$ 0.5	6.2 $\pm$ 0.7
2	Chymotrypsin	4.5 $\pm$ 0.9	16.8 $\pm$ 1.0
1	Chymotrypsin	0.5 $\pm$ 0.2	9.8 $\pm$ 1.2
1	Tryptic extract	15.8 $\pm$ 1.0	15.3 $\pm$ 1.1
		11.6 $\pm$ 0.8	

TABLE 5  
AUXIN YIELD FROM CHLOROPLASTIC AND CYTOPLASMIC PROTEINS COMPARED ON A PER-MILLIGRAM-OF-NITROGEN BASIS

NO. OF 10-MG. DIGESTIONS	ENZYME	AVERAGE <i>Avena</i> CURVATURE IN DEGREES	
		CHLOROPLASTIC PROTEIN	CYTOPLASMIC PROTEIN
2	Control	3.6	3.5
1	Control	0	0
1	Trypsin	5.5	6.2
2	Chymotrypsin	6.3	16.8
1	Chymotrypsin	0.7	9.8
1	Tryptic extract	19.2*	15.3

\* Average of two determinations.

*Diffusion Experiments.*—The question can be raised whether the auxin released from the proteins is similar to the auxin found in the leaf. Corresponding diffusion rates of the two auxins should indicate a similarity in

molecular weights. The diffusion technique has been described by a number of workers;<sup>8, 15</sup> accordingly, experiments were performed in general following previous procedures comparing the auxin obtained from ether extracts of spinach leaves to the auxin liberated by enzymatic hydrolysis of the cytoplasmic protein. At the same time, these auxins were compared with a growth substance of known constitution, crystalline indole acetic acid. Agar blocks containing the auxin preparations, purified by

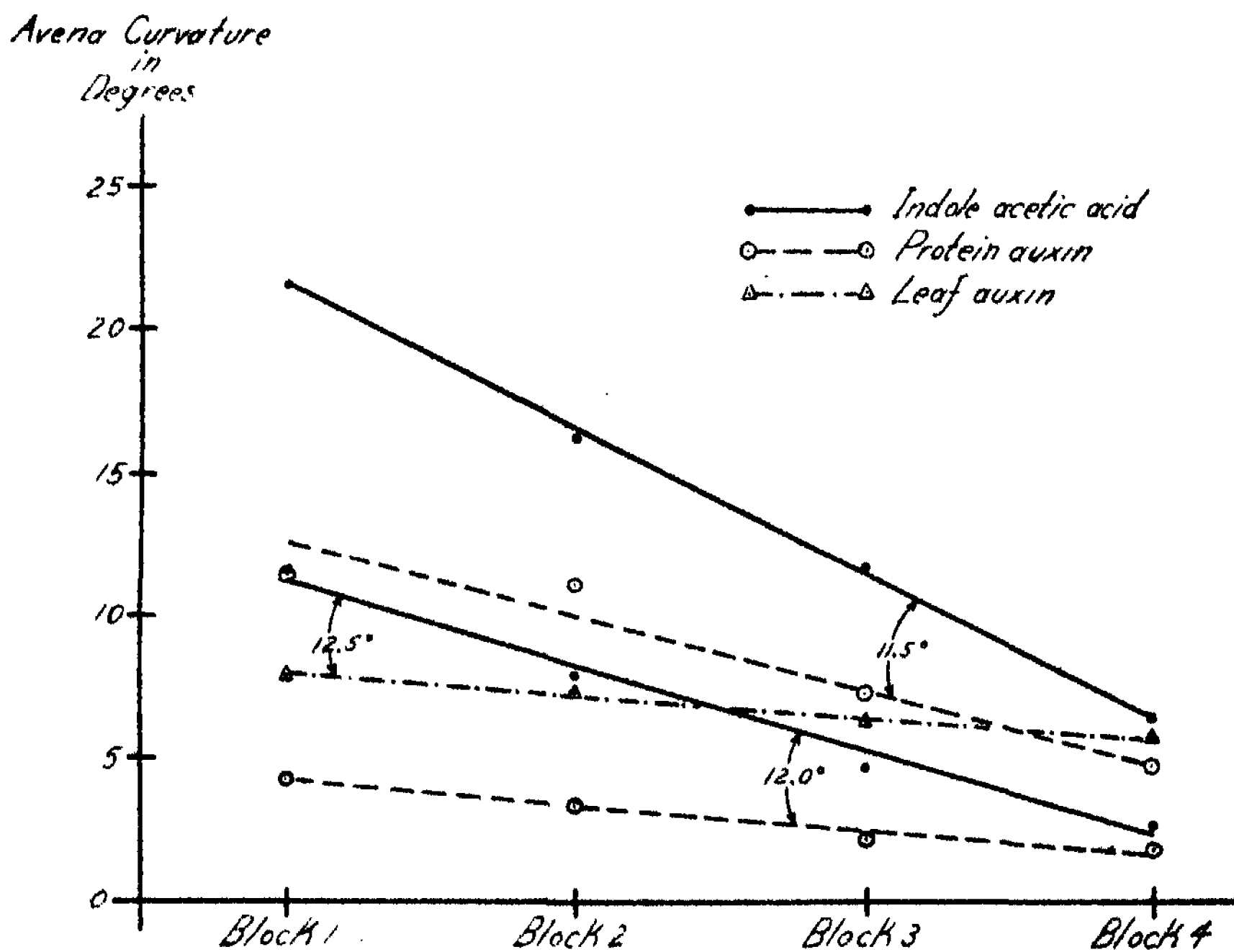


FIGURE 1

The distribution of auxin in four agar blocks, each 1.6 mm. thick, after 90 minutes' diffusion. The two upper curves represent one experiment, while the three lower curves are from another experiment. The initial concentration of indole acetic acid was 100 gammas per liter.

the donor-block method suggested by van Overbeek,<sup>12</sup> were placed upon three blocks of pure agar, each 1.6 mm. thick, and diffusion was allowed to ensue for 90 minutes.

If the concentrations of auxin in the four blocks are within the proportionality range, the resulting *Avena* curvatures plot graphically as a straight line. The slope of this line is dependent upon the time of diffusion, the temperature and the weight of the auxin molecule. If the temperature and time of diffusion are held constant for the auxins under comparison in a

given experiment, the slopes should vary as a function of the molecular weights. Hence, we have used contrasting slopes as a criterion of dissimilar molecular weights.

In figure 1 are shown data from two diffusion experiments. The angular difference between the indole acetic acid and enzyme auxin curves of one experiment is  $11.5^\circ$ ; in the other experiment, it is  $12.0^\circ$ , and the angular difference between the leaf auxin and indole acetic acid curves is  $12.5^\circ$ . Parallel data have been obtained from repeated experiments. There is a dissimilarity between the slopes of the curves obtained with indole acetic acid and enzyme auxin, and a similarity between the slopes of the curves obtained with enzyme and leaf auxins.

*Discussion and Interpretation of Experimental Results.*—The preceding data establish that an *Avena* growth-accelerating substance is definitely associated with proteins isolated from the leaves of spinach; the auxin can be released from the proteins through the action of various proteolytic enzymes. Tryptic extract appears to yield a greater amount of auxin than crystalline trypsin or chymotrypsin. This result would indicate that the absolute amount of auxin released depends upon the extent of hydrolysis of the proteins, if it is assumed that tryptic extract contains a mixture of at least three proteolytic enzymes, trypsin, chymotrypsin and carboxypolypeptidase<sup>2</sup> and hence, will split a protein more completely than the crystalline enzymes. However, it is not argued that the enzymes used in this study represent those present in the plant. On the contrary, we consider it more plausible that the plant contains other enzymes which function to release auxin from proteins. These results explain at least in part the hydrolytic release of auxin from plant material reported by Gustafson,<sup>7</sup> Link, Eggers and Moulton<sup>10</sup> and others. The elucidation of the position that auxin occupies in the protein or in association with the proteins must, of course, await further experimentation.

The possibility that association of auxin with proteins is an adsorption phenomenon should be considered. We feel that the evidence weighs against a hypothesis of purely physical adsorption. In the first place, auxin released from proteins by hydrolysis is organophilic; hence, if it were adsorbed on proteins, prolonged extraction with organic solvents should elute the auxin from the protein. No auxin has been obtained by extraction of the proteins with ether and ether-alcohol. Second, liquefaction of proteins by enzymes is complete within twelve hours, but auxin cannot be obtained consistently until the proteins have been hydrolyzed for 72 hours. If the process is one of adsorption, it might reasonably be assumed that auxin would be released as soon as the protein had been liquefied.

It seems unlikely that substances of low molecular weight such as amides, peptides, etc., should be contained in these protein preparations. The washings after cytolysis of the leaf to remove cell fluids, together with



the narrow pH range of cytoplasmic protein precipitation, and extensive washing of the precipitates throughout the process of preparation would tend to remove water soluble impurities.

We hold that our experiments are valid for spinach proteins and are verifiable on the same material. Although one might extend these results to other leaf material on the basis of Chibnall's work which showed that the differences were "strikingly small" between the leaf proteins prepared from five separate plant families, one would certainly not be justified without extensive further work in drawing an analogy between these proteins and those contained in seeds, fruits, etc.

The consideration that auxin is found in both the chloroplastic and cytoplasmic proteins leads to an interesting speculation: does auxin function in the synthesis of these proteins, or is auxin represented as an end-product of protein metabolism? If the former hypothesis can be substantiated, then the presence of auxin in both fractions can perhaps be explained. If the latter is found to be the true explanation, why is auxin not confined to the cytoplasm which is in contact with the cell wall upon which its function has been explained, according to Went and Thimann, by Heyn?

It has been shown that the auxin content of one fraction of cytoplasmic protein is greater than in another when the two fractions have been separated from each other by isoelectric precipitation. This fact at once suggests the desirability of separating and crystallizing these proteins. Not until this has been accomplished can an adequate study of the amount of auxin released in relation to the extent of protein hydrolysis be undertaken, since at the present stage it is impossible to tell if auxin is associated with some proteins, but not with others, or associated with all proteins but in varying amounts. Chibnall believes that protein preparations of leaves "are really mixtures of many proteins which have similar solubility properties" (p. 141), although they can be conveniently classified as glutelins on the basis of their insolubility in water and solubility in a slight excess of either acid or base. It may also be possible to isolate the enzymes responsible for auxin release in the plant since they are apparently present in some of the protein preparations.

The identification of the auxin released by enzymatic hydrolysis of the leaf protein with the free auxin of the leaf can only be determined by actual isolation of the auxins. From the data presented in figure 1, and other experiments, it appears that the auxin released from proteins is of similar molecular size to the free auxin of the leaf. The dissimilarity between the slopes of the curves for enzyme and leaf auxin, on the one hand, as contrasted with those for indole acetic acid on the other suggests that the auxin in the leaf and the auxin released from proteins are of a lower molecular weight than indole acetic acid.



*Summary.*—1. Auxin has been shown to be associated with proteins isolated from the leaves of spinach. It can be released by enzymatic hydrolysis.

2. Auxin is obtained from both the cytoplasmic and chloroplastic proteins by enzymatic hydrolysis of the proteins. The cytoplasmic proteins have been separated into two fractions by isoelectric precipitation; more auxin is obtained from one of these fractions than from the other.

3. Diffusion experiments indicate the similarity between the leaf auxin and the auxin released from proteins by enzymatic hydrolysis. Both auxins appear to be of lower molecular weight than indole acetic acid.

4. Evidence is presented that the auxin released from leaf proteins of spinach is not a result of bacterial contamination.

\* We gratefully thank Dr. F. G. Gustafson and Professor H. H. Bartlett, Department of Botany, University of Michigan, for their aid, suggestions, criticisms and use of facilities. This work was performed while the authors held F. C. and Susan Eastman Newcombe Fellowships in plant physiology.

† Papers from the Department of Botany, University of Michigan, No. 807.

‡ We are indebted to Dr. Walter S. Hales of the Enzyme Research Laboratory, Bureau of Agricultural Chemistry and Engineering, U. S. Department of Agriculture, for providing us with the crystalline papain used in this study.

<sup>1</sup> Avery, G. S., Jr., Berger, J., and Shalucha, B., *Amer. Jour. Bot.*, **28**, 596 (1941).

<sup>2</sup> Bodansky, M., *Introduction to Physiological Chemistry*, New York, 1938.

<sup>3</sup> Chibnall, A., *Protein Metabolism in the Plant*, New Haven, 1939.

<sup>4</sup> Gortner, R., *Outlines of Biochemistry*, New York, 1938.

<sup>5</sup> Granick, S., *Amer. Jour. Bot.*, **25**, 558 (1938).

<sup>6</sup> Granick, S., *Ibid.*, **25**, 561 (1938).

<sup>7</sup> Gustafson, F., *Science*, **92**, 266 (1940).

<sup>8</sup> Heyn, A., *Proc. Kon. Akad. Wetensch. Amsterdam*, **38**, 1074 (1935).

<sup>9</sup> Kögl, F., Erxleben, H., and Haagen Smit, A., *Zeit. physiol. Chem.*, **225**, 215 (1934).

<sup>10</sup> Link, G., Eggers, V., and Moulton, J., *Bot. Gaz.*, **102**, 590 (1941).

<sup>11</sup> Menke, W., *Zeit. Bot.*, **32**, 273 (1938).

<sup>12</sup> Overbeek, J. van, *Plant Physiol.*, **15**, 291 (1940).

<sup>13</sup> Skoog, F., and Thimann, K., *Science*, **92**, 64 (1940).

<sup>14</sup> Tauber, H., *Enzyme Chemistry*, New York, 1937.

<sup>15</sup> Went, F., *Rec. trav. bot. néerl.*, **25**, 1 (1928).

<sup>16</sup> Went, F., and Thimann, K., *Phytohormones*, New York, 1937.

## THE DISTRIBUTION OF X-RAY INDUCED CHROMOSOMAL ABERRATIONS

By KARL SAX

ARNOLD ARBORETUM, HARVARD UNIVERSITY\*

Communicated April 30, 1942

The distribution of chromosomal aberrations in *Tradescantia* microspores is not at random for the various loci of the chromosome arms. Both spontaneous (Giles<sup>1</sup>) and x-ray induced (Sax and Mather<sup>2</sup>) aberrations occur more frequently near the centromere than in the distal loci of the chromosome arms. This distribution is found for both simple terminal deletions and for the more complex aberrations involving fusions or exchanges between different chromosomes. An excess of x-ray induced breaks in the proximal regions of the chromosome arms have been found in *Crepis* by Lewitsky and Sizova,<sup>3</sup> and in the grasshopper *Circotettix* by Helwig.<sup>4</sup> The distribution of breaks in *Drosophila* chromosomes is more nearly at random (Kaufmann<sup>5</sup>), but here the breaks produced by raying the sperm do not result in new associations until fertilization of the egg several days later. Many types of aberrations are eliminated, due to their lethal effect, before the salivary gland chromosomes are analyzed.

The actual or potential breaks produced in the chromosomes by x-rays must be at random, but the production of chromosomal aberrations resulting from deletions or from illegitimate fusions between broken ends of chromosomes is determined, to a considerable extent, by secondary factors. One of these factors is the spatial relations of the chromosomes. One break in a single chromosome usually is followed by restitution; but, if the chromosome is split into two sister chromatids, a break may be followed by lateral fusion of broken ends of the two chromatids to produce a dicentric chromosome and a deletion. Fusions between broken ends of different chromosomes occur only when the breaks are in close proximity. The absence of free recombinations is shown by the proportions of the various types of chromosomal aberrations. In the salivary gland chromosomes of *Drosophila* the induced inversions, in relation to reciprocal translocations, are twice as frequent as expected.<sup>5</sup> In *Tradescantia* the proportion of ring chromosomes, as compared with dicentrics, is about three times that expected with free recombination of broken ends of chromosomes. It is evident that propinquity favors aberrations within a single chromosome.

Another factor in the production of effective breaks in chromosomes is the rôle of the centromere in relation to the coiling cycle and polarity of the chromosomes. Acentric chromosome fragments are less sensitive to x-rays than are the normal chromosomes.<sup>6</sup> When breaks occur in adjacent

arms of different chromosomes the illegitimate fusion may result in either a translocation or a dicentric chromosome. The analysis of such aberrations induced at prophase shows that dicentrics are twice as frequent as translocations.<sup>7</sup> This non-random distribution indicates that lateral fusions are more frequent than diagonal fusions in the polarized chromosomes, and suggests that the coiling mechanism of the chromosomes at this stage of nuclear development may impose stresses which separate the chromosome at the point of breakage. The analyses of differential sensitivity of chromosomes to x-rays<sup>8</sup> also support the assumption that chromosome movement is a potent secondary factor in the production of chromosomal aberrations. Chromosome orientation and chromosome movement are, to a considerable extent, under the control of the centromere.

The relative sensitivity of centric and acentric chromosomes to x-ray has been determined by analyzing the results of successive exposures. The microspores were rayed during the resting stage to produce chromosome rings and dicentrics and their accompanying acentric fragments. The same microspores were irradiated two days later at prophase, and the cells were fixed and stained a day later. The second exposure produced only chromatid aberrations which could be distinguished from the chromosome aberrations induced by the first exposure. Of the 11,502 chromosomes examined there were 681 acentric chromosome segments released by the production of ring and dicentric chromosomes. These acentric fragments were equivalent in length to 340 normal chromosomes or 3.0 per cent of the total chromosome complement. Of the 1215 chromatid breaks only 4 or 0.3 per cent were found in acentric chromosome fragments. Thus the breaks in acentric chromosomes were only about a tenth as frequent as expected if aberrations were produced at random.

The low incidence of x-ray induced aberrations in acentric chromosome fragments may be related to the absence of a centromere, or it may be associated with the low incidence of multiple aberrations in a single chromosome, either centric or acentric. The second possibility was tested by analyzing the relative frequency of the various types of aberrations induced at early prophase. The *Tradescantia* microspores were given 320 r and the cells were fixed 25-30 hours after raying. Among the 3840 chromosomes examined there were 715 chromosomes with a 1-hit chromatid aberration and 266 with a 2-hit chromatid aberration, a total of 25.6 per cent. There were also 502 chromosomes with 2-hit chromosome aberrations, or 13.8 per cent. Only 8 chromosomes were found with both a chromatid and a chromosome aberration in the same chromosome. If aberrations were at random we should expect 3.5 per cent of the chromosomes to have such multiple aberrations, but only 0.2 per cent were observed. The paucity of such multiple aberrations suggests that one effective break in a chromosome inhibits the occurrence of a second effec-

tive break, but since we are dealing with combinations of chromatid and chromosome aberrations a short period of transition from one type to the other would also account for the low frequency of these multiple aberrations.

TABLE 1  
FREQUENCY OF CHROMOSOMAL ABERRATIONS. TOTAL CHROMOSOMES 2334

TYPE OF ABERRATION	PER CENT OF CHROMOSOMES	
	OBSERVED	EXPECTED
One 1-hit chromatid	11.8 %	
One 2-hit chromatid	15.0 %	
One 2-hit chromosome	9.4 %	
Multiple aberrations:		
Two 1-hit tid	0.04%	1.4%
Two 2-hit tid	0.7 %	2.2%
Two 2-hit chromosome	0.0 %	0.9%
1-hit + 2-hit tid	0.3 %	1.8%
1-hit tid + 2-hit chromosome	0.04%	1.1%
2-hit tid + 2-hit chromosome	0.2 %	1.4%
Total multiple aberrations	1.28%	8.8%

A second analysis of chromatid and chromosome aberrations was made in which all of the various types of multiple aberrations were included. The cells were fixed at 30-32 hours after irradiation of the microspores. The results of the analysis are shown in table 1. The expected frequency of multiple aberrations in individual chromosomes is taken as the product of the frequencies of the single events. It is evident that multiple aberrations are much less frequent than expected. The low incidence of both the multiple aberrations of similar types, and the multiple aberrations involving both chromosome and chromatid alterations, indicate that random union of broken ends does not occur.

An analysis of multiple aberrations involving alterations induced during the resting stage also indicates that fusions of broken ends of chromosomes are not at random. The data are shown in table 2. In the 1000 cells examined there were 244 centric ring chromosomes and 684 dicentric chromosomes, a ratio of 1 : 2.8 instead of the 1 : 10 ratio expected if fusions of broken chromosome ends were at random. Only the dicentrics have been included in the subsequent analysis of multiple aberrations. Of the 6000 chromosomes analyzed 1270 or 21.2 per cent had only one break, and 49 or 0.8 per cent had two breaks which led to the formation of multicentrics. Random production and fusion of effective breaks should have produced two effective breaks in each of 276 chromosomes or 4.6 per cent. The deficiency of multiple chromosome breaks in the same chromosome is of the same order as the deficiency of multiple aberrations involving either

chromatid or chromosome breaks (table 1). The frequency and types of multicentric chromosomes per cell are more nearly in accord with theoretical expectations. We may assume that fusions can occur between any two of the twelve chromosome arms. With two dicentric bridges per cell we would expect the following combinations: 24 cells with 2 dicentrics, 16 cells with a tricentric chromosome and 1 with a closed dicentric. With 3 dicentric bridges there should be cells with 3 dicentrics, 1 dicentric plus a tricentric, 1 tetracentric, or a closed tricentric, in the ratio of 2 : 12 : 6 : 1. The observed and expected frequencies are shown in table 2. In cells with 2 dicentric bridges there is an excess of dicentrics at the expense of tricentric chromosomes. The deficiency of tricentrics might be expected since one of the three chromosomes involved in the production of a tricentric was effectively broken in each of its two arms, and the previous data show that multiple breaks in the same chromosome occur much less frequently than expected. In cells with three dicentrics the types of aberrations occur with the expected frequency. Apparently with a greater number of effective breaks a more nearly random fusion of broken ends is possible.

The total number of multicentric aberrations per cell also approaches the theoretical frequency. Here the maximum number of aberrations is limited by the number of chromosome arms which may be broken and rejoined in new associations. The calculation of expected frequencies has been done by Dr. Reed. The number of dicentric bridges expected per cell is the total number of cells times the appropriate term of the expansion of  $[(1 - q) + q]^n$ , where  $q$  is the frequency of chromosome arms involved in dicentric aberrations and  $n$  is the pairs of chromosome arms. The observed and theoretical frequencies of dicentric bridges are shown in table 2. Since the  $\chi^2$  test gives a value of  $P = 0.02$  it is probable that the deficiency of two or more dicentric aberrations per cell is significant, but the deviation is small compared with deficiencies of expected chromosomes with multiple breaks (table 1).

TABLE 2  
FREQUENCY AND TYPES OF MULTICENTRIC CHROMOSOMES PER CELL. DOSE 500 R AT  
70 R/M

DICENTRIC BRIDGES	TYPES	NUMBER OBSERVED	FREQUENCY EXPECTED	TOTALS OBSERVED	EXPECTED
0	Cell normal	457		457	483
1	1 dicentric	418		418	373
2	{ 2 dicentric	81	68 }	110	120
	{ 1 tricentric	29	44 }		
	{ 3 dicentric	1	1 }	14	21
3	{ 1 di- + 1 tricentric	9	8 }		
	{ 1 tetracentric	4	4 }		
4	1 pentacentric	1		1	3

The great deficiency of multiple breaks in a single chromosome of a multicentric complex can be attributed to the inhibiting effects of the first break or to the spatial relations of the chromosomes. If an effective break occurs in a chromosome arm, certain stresses imposed by polarity may be released which inhibit further effective breaks in the same arm. Such an inhibiting mechanism may account, in part, for the low frequency of chromatid aberrations in acentric chromosome fragments. This mechanism would not be expected to operate across the centromere and affect the incidence of effective breaks in the other arm of the same chromosome. This assumption is supported by the fact that ring chromosomes occur much more frequently than expected. The high frequency of ring chromosomes must be attributed to the relatively close association of the two arms of the same chromosome. The low incidence of multiple breaks in a single chromosome involved in multicentric associations can be attributed in part to the spatial relations which prevent simultaneous fusions involving both arms of the same chromosome.

*Summary.*—The relative frequencies of various types of chromosomal aberrations indicate that secondary factors play an important part in determining the frequency of effective breaks induced by x-rays. These secondary factors include: (1) the spatial relations of the chromosomes, (2) the stresses imposed by the centromere in maintaining polarity and (3) the stage of chromosome development in the nuclear cycle.

\* This work was supported, in part, by a grant from the International Cancer Research Foundation.

<sup>1</sup> Giles, N., *Proc. Nat. Acad. Sci.*, 26, 567-575 (1940).

<sup>2</sup> Sax, K., and Mather, K., *Jour. Genet.*, 37, 483-490 (1939).

<sup>3</sup> Lewitsky, G., and Sizova, M., *C. R. Acad. Sci. U.S.S.R.*, 4, 1-70 (1935).

<sup>4</sup> Helwig, E. R., *Jour. Morph.*, 55, 265-297 (1933).

<sup>5</sup> Kaufmann, B. P., *Cold Spring Harbor Symposia on Quant. Biol.*, 9, 82-92 (1941).

<sup>6</sup> Sax, K., *Jour. Gen. Physiol.*, 25, 533-537 (1942).

<sup>7</sup> Sax, K., *Genetics*, 23, 494-516 (1938).

<sup>8</sup> Sax, K., and Swanson, C. P., *Amer. Jour. Bot.*, 28, 52-59 (1941).

GENETIC CONTROL OF BIOCHEMICAL REACTIONS IN *NEURO-SPORA*: AN "AMINO BENZOICLESS" MUTANT\*

By E. L. TATUM AND G. W. BEADLE

DEPARTMENT OF BIOLOGY, STANFORD UNIVERSITY

Communicated April 17, 1942

Para-aminobenzoic acid has recently been recognized as a factor required for the growth of a number of microorganisms<sup>1</sup> and as a member of the vitamin B group.<sup>2</sup> One of the number of x-ray induced mutants of *Neurospora crassa*, obtained as described elsewhere,<sup>3</sup> is characterized by the loss of ability to synthesize *p*-aminobenzoic acid. This "aminobenzoicless" mutant is differentiated from normal by a single gene, is unable to grow on unsupplemented synthetic medium, but its growth is indistinguishable from normal when *p*-aminobenzoic acid is supplied.

The mutant culture was maintained on "complete" medium containing yeast extract, malt extract and sucrose. Conidia from this stock culture were used for inoculating test cultures. The test medium was the "synthetic" or "minimal" medium previously described<sup>3</sup> containing inorganic salts and nitrogen, sucrose and 4 gammas of biotin<sup>4</sup> per 1000 ml. Agar used in this synthetic medium was freed of *p*-aminobenzoic acid by repeated washing with distilled water. Cultures were incubated at 25°C.

Crosses were made of the *p*-aminobenzoicless mutant (sex *A*, the equivalent of + in Lindegren's terminology<sup>6</sup>) with the normal strain (sex *a*); ascospores were dissected in order from asci and grown on complete medium. These cultures were tested by transferring them to minimal medium and observing for growth. In all, 142 single spore cultures were grown and tested. Of these, 72 proved to be able to synthesize *p*-aminobenzoic acid and 70 were unable to accomplish this. Of 11 asci from each of which all eight ascospore cultures were obtained, all had four normal and four mutant-type spores, and in no case were more than four spores of a kind obtained from any one ascus. The results in the 17 asci in which segregations for both sex and aminobenzoicless could be deduced are summarized in table 1. It is evident from these results that the aminobenzoicless character is inherited as though it were differentiated from normal by a single gene. There is no evidence of linkage of aminobenzoicless and sex. The fact that in 10 out of 17 asci aminobenzoicless segregated from its normal allele in the second division indicates that the aminobenzoicless locus is an appreciable distance from the centromere.

In addition to the original aminobenzoicless mutant (ascospore isolate 1633), which was used for all physiological investigations described except those otherwise noted, an apparently independent occurrence of this mutant was recorded (isolate 5359). A cross between 5359 and 1633-63-6



TABLE 1

SEGREGATION FOR SEX AND ABILITY TO SYNTHESIZE *p*-AMINO BENZOIC ACID, IN ASCI FROM A CROSS OF NORMAL (SEX *a*) AND AMINO BENZOICLESS (SEX *A*)

The base and apex of the ascus are regarded as equivalent and the orientation of second division spindles is disregarded unless this has significance with regard to recombination. Since the third division is equational, each pair of products of this division is reported as a unit.

TYPE OF SPORE ARRANGEMENT				NUMBER OF ASCI OBSERVED
A +	A +	<i>a pab</i>	<i>a pab</i>	3
<i>A pab</i>	<i>A pab</i>	<i>a</i> +	<i>a</i> +	4
A +	<i>A pab</i>	<i>a</i> +	<i>a pab</i>	9*
A +	<i>a</i> +	<i>A pab</i>	<i>a pab</i>	0
A +	<i>a pab</i>	A +	<i>a pab</i>	0
<i>A pab</i>	<i>a</i> +	<i>A pab</i>	<i>a</i> +	0
A +	<i>a pab</i>	<i>A pab</i>	<i>a</i> +	1

\* In two asci the results were such that it was presumed that two spores had been transposed in dissection.

TABLE 2

RESPONSE OF *p*-AMINO BENZOICLESS MUTANT 1633 TO COMPOUNDS RELATED TO *p*-AMINO BENZOIC ACID

*Inactive Compounds*

Benzoic acid	Sulfanilic acid
<i>o</i> -Aminobenzoic acid	Tyrosine
<i>m</i> -Aminobenzoic acid	2, 4-Diaminotoluene
<i>p</i> -Hydroxybenzoic acid	Acetanilide
<i>p</i> -Chlorobenzoic acid	<i>p</i> -Methylcyclohexanol

*Active Compounds*

SUBSTANCE	ACTIVITY COMPARED WITH THAT OF <i>p</i> -AMINO BENZOIC ACID*	
	BY WEIGHT IN LIQUID MEDIUM	BY GROWTH RATE ON SOLID MEDIUM
Acetyl- <i>p</i> -aminobenzoic acid	1/50†	
<i>p</i> -Nitrobenzoic acid	1/100	1/120
<i>p</i> -Dimethylaminobenzaldehyde	1/1150‡	1/170‡
<i>p</i> -Toluidine	1/1400	1/1600
Aniline	1/13,000	
<i>p</i> -Nitrotoluene§	1/14,000	1/5000
<i>p</i> -Aminoacetophenone§	1/100,000	

\* Activity values based on several series of varied concentrations, and when possible calculated from the concentrations giving 1/3 maximum growth.

† Value obtained by visual comparison with control *p*-aminobenzoic dilutions.

‡ The difference between these two values seems to be correlated with the use of liquid and solid media.

§ Only very slight growth possible with any concentration, figures for activity therefore not too significant.

(derived mutant of sex *a*) gave 68 offspring from single ascospores taken at random, and all of these failed to grow in minimal medium. Presumably



these two aminobenzoicless strains represent independent occurrences of the same gene change. It is improbable that the two strains are the result of mutations of closely linked genes or that the second represents merely a tube contaminated with a spore of the first, although these two possibilities cannot be rigorously excluded.

Physiologically the two aminobenzoic strains are similar if not identical. Although exhaustive measurements of the *p*-aminobenzoic acid require-

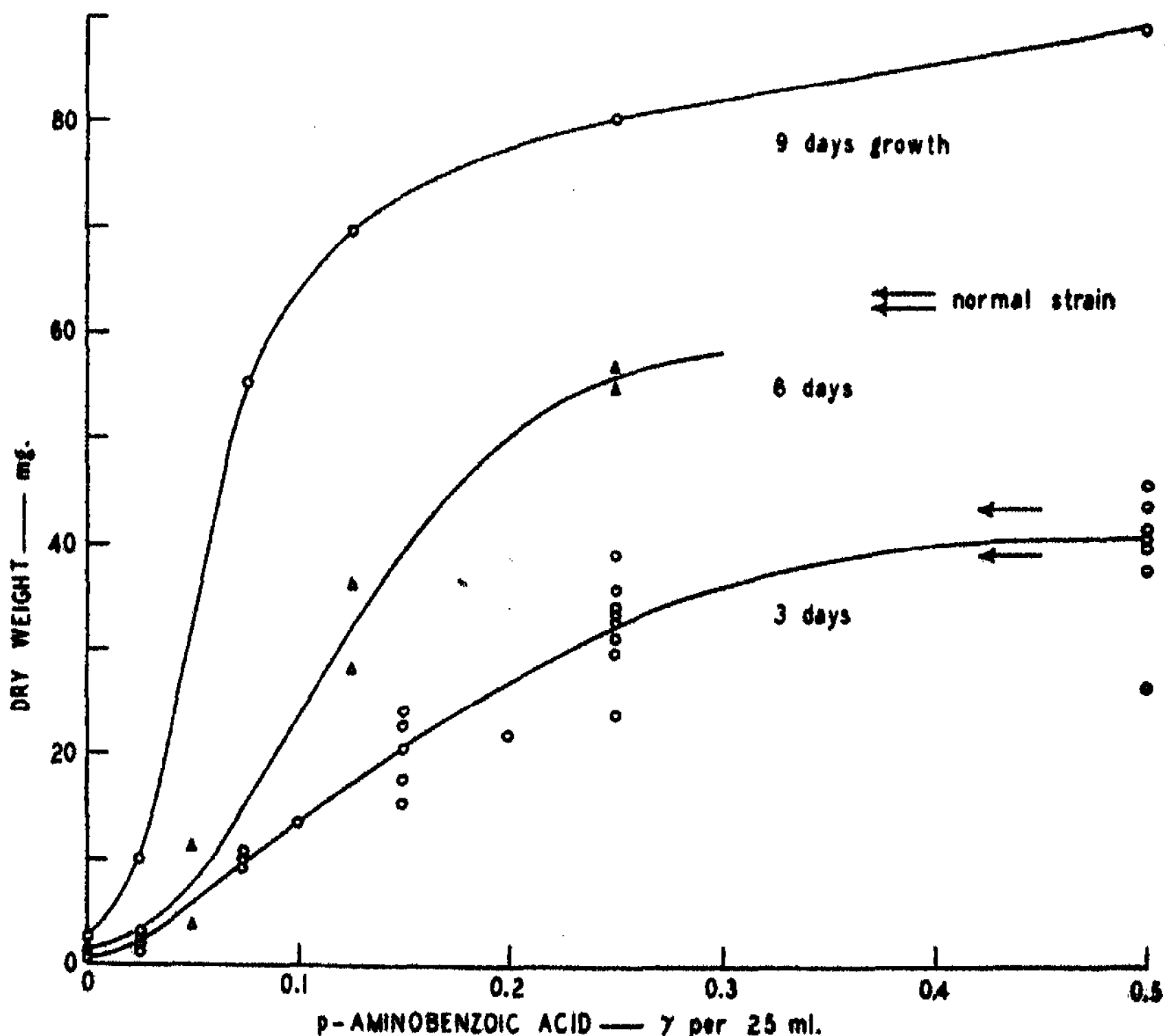


FIGURE 1

Dry weight attained after three, six and nine days as a function of *p*-aminobenzoic acid concentration. Weights of normal strain indicated by arrows.

ment of strain 5359 have not been made, it can be said that this is of the same order of magnitude as that of strain 1633. The quantitative growth responses to the compounds related to *p*-aminobenzoic acid which are listed in table 2 are similar in the two strains. Furthermore they are similarly inhibited by sulfanilamide as indicated in figure 4.

Growth of the aminobenzoicless mutant is normal only in the presence of *p*-aminobenzoic acid. In tests of other known growth factors a slight

but significant response to pimelic acid has been noted. This response is apparent only after a period of five or six days, whereas that to *p*-aminobenzoic acid is evident after one day. The pimelic acid effect has been observed with both liquid and solid media and with carefully recrystallized pimelic acid. This phenomenon, which obviously needs further investigation, suggests either some sort of adaptation or a possible conversion of pimelic acid to *p*-aminobenzoic acid.<sup>6</sup> Whatever its nature may be,

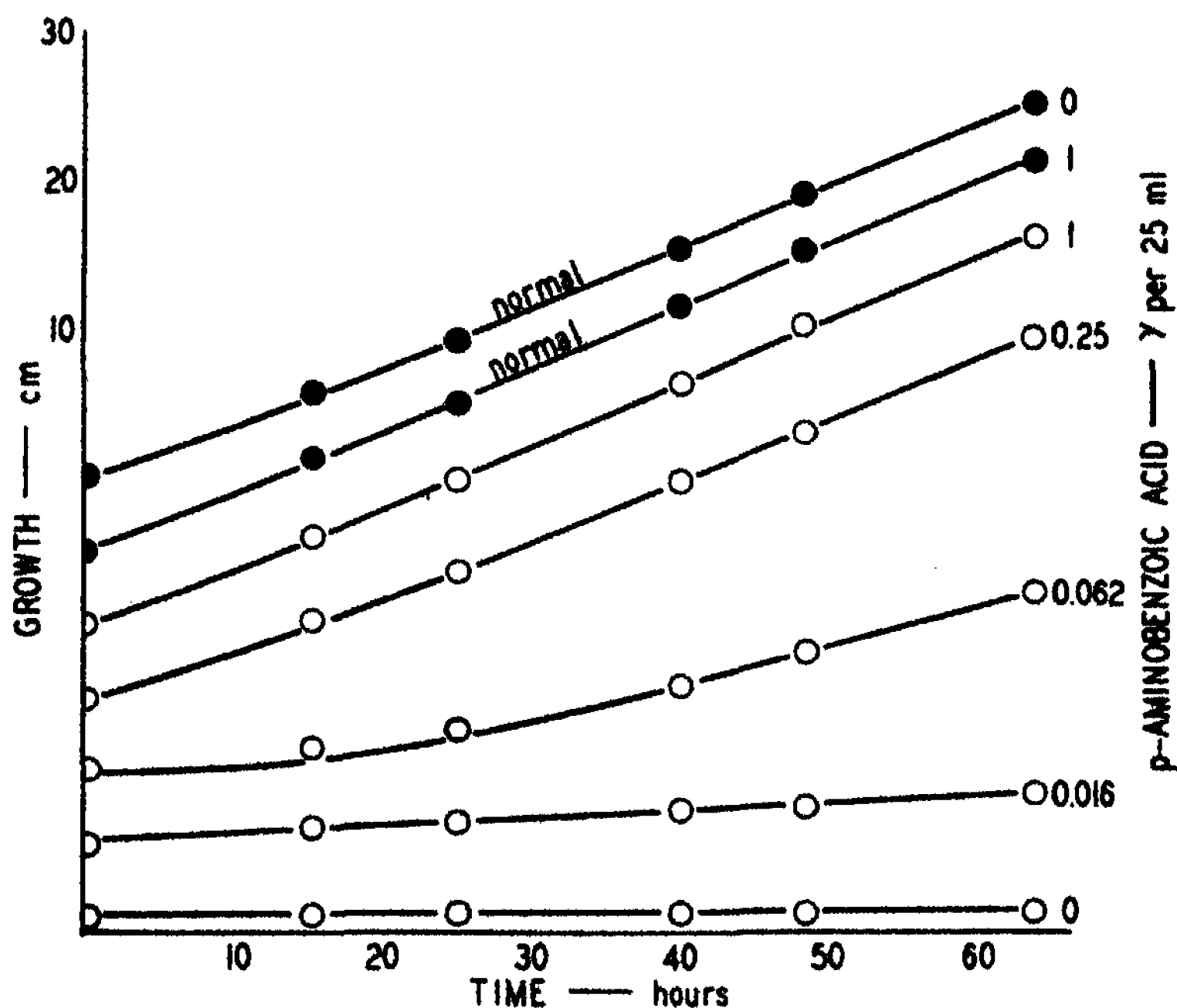


FIGURE 2

Position of mycelial frontiers in growth tubes as a function of time. Results with normal strains indicated by solid circles and those with *p*-aminobenzoicless strain (1633) indicated with open circles. Vertical scale offset for each successive curve in the series.

recovering and retesting strains grown in the presence of pimelic acid shows that the effect does not involve a genetic change in the organism.

The growth of the aminobenzoicless mutant is a function of the amount of *p*-aminobenzoic acid supplied to it. This quantitative relation can be determined either by the increase in dry weight of mycelium in liquid medium, or by rate of progression of the mycelium front through a tube partly filled with agar medium.

For the dry weight determinations, cultures were grown in 250-ml.

Erlenmeyer flasks containing 25 ml. of synthetic medium supplemented with *p*-aminobenzoic acid. Each flask was inoculated with a drop of a water suspension of conidia. At the end of the incubation period, generally 3 days, the mycelium was removed, washed and dried at 105°C. on a weighed watch glass. Figure 1 gives the results of a number of series. The relation between weight and *p*-aminobenzoic acid is almost linear up to a concentration of 0.25 gamma per 25 ml. The average weight reached

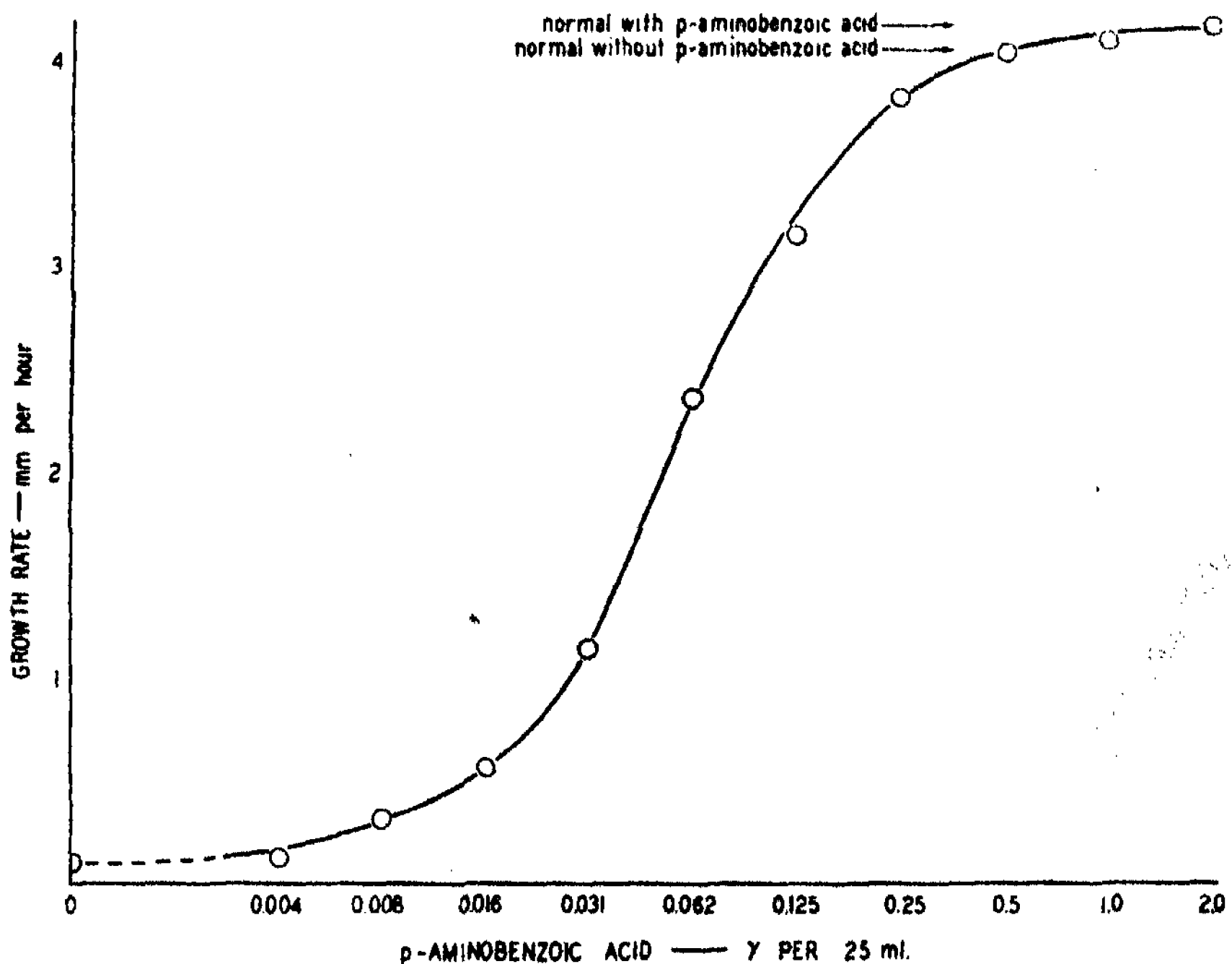


FIGURE 3

Rate of growth in tubes of aminobenzoicless strain (1633) as a function of *p*-aminobenzoic acid content of medium. Rates of normal strains indicated by arrows. Individual points obtained by averaging four independently obtained values.

in 3 days with 0.5 gamma per 25 ml. is about 40 mg., the weight reached by the normal, unsupplemented culture in the same time. The variability is rather great, due to the fact that only one determination can be made for each culture and due to variations in inoculum, length of initial lag period and probably other factors. Because of these difficulties and because significant weights cannot be reached in much less than 3 days, this method is only fairly satisfactory for assay purposes.

A more satisfactory method of determining the relation between *p*-aminobenzoic acid concentration and growth is to measure the rate of

progression of the mycelium front along the surface of agar medium in a horizontal tube as previously described.<sup>3</sup> This method has the advantage that variations inherent in the liquid culture method are not included in the measurements, since the final rate of progression is determined after equilibrium is reached and several measurements may be made in a given culture tube. This method has the further advantage that significant rate measurements may be made over a relatively short period of time, after the rate has become constant. The results of a series of *p*-aminobenzoic acid concentrations are shown in figure 2. As more *p*-aminoben-

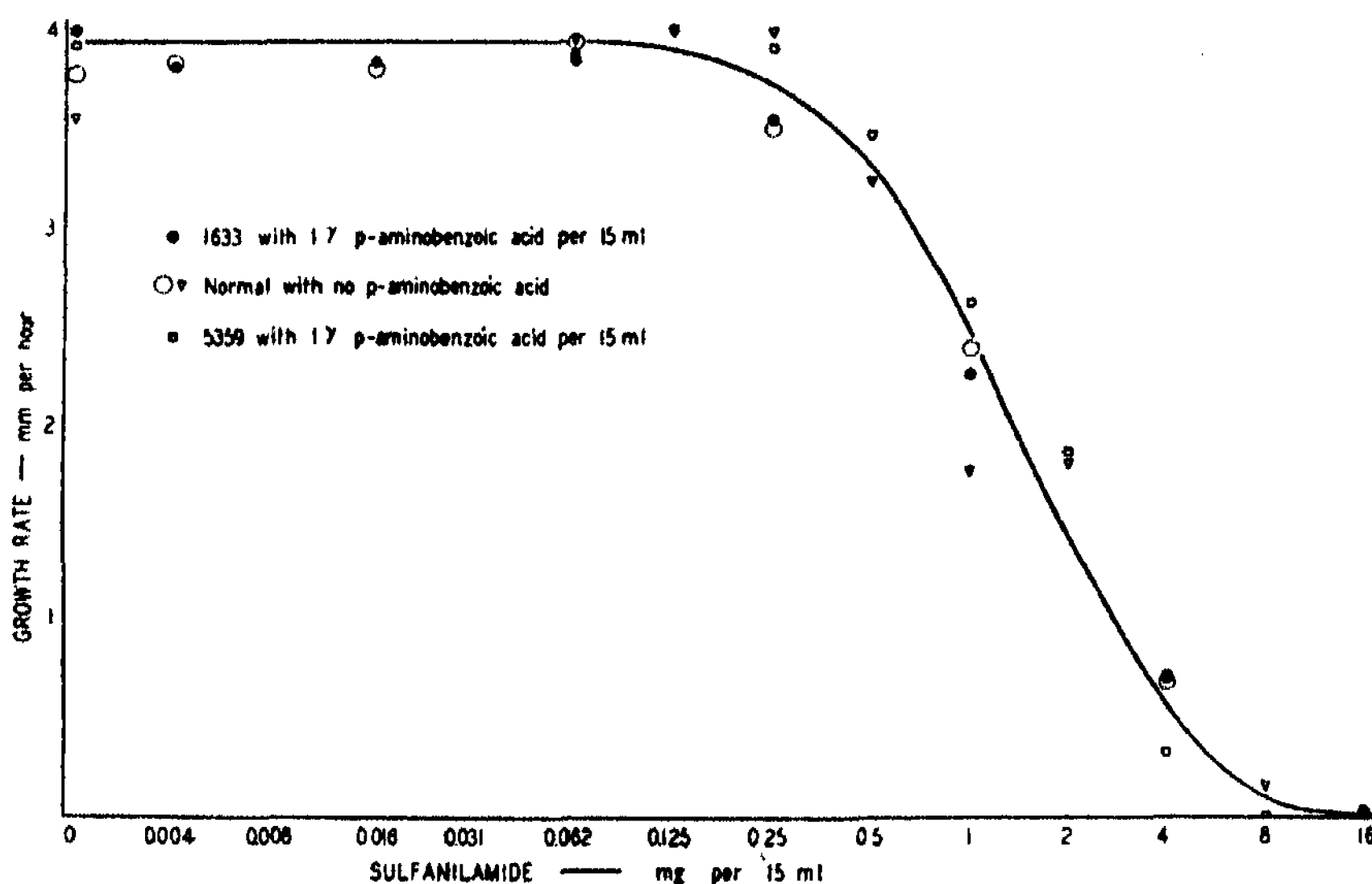


FIGURE 4

Inhibition of normal and aminobenzoicless strains (1633 and 5359) with sulfanilamide. Mutant strains supplied with one gamma *p*-aminobenzoic acid per 15 ml. medium.

zoic acid is supplied the rate of progression increases until with 0.5 gamma per 25 ml. it is the same as that of the normal strain (approximately 4 mm. per hour at 25°C. Figure 3 gives the final relation between the rate and vitamin concentration.

The results of both methods show that when enough *p*-aminobenzoic acid is supplied to the mutant it is indistinguishable from the normal strain. The inability to synthesize this vitamin seems therefore to be the only differentiating factor.

The biological significance of *p*-aminobenzoic acid was first recognized as a result of its action in overcoming the bacteriostatic or inhibitory

effects of sulfanilamide.<sup>7</sup> This antagonism has been interpreted as a competitive reaction, with an excess of sulfanilamide displacing the normally present *p*-aminobenzoic acid from its functional rôle, and an excess of *p*-aminobenzoic acid overcoming the resulting inhibition. The widespread existence of this *p*-aminobenzoic acid—sulfanilamide antagonism has been taken as indicating the essential nature of *p*-aminobenzoic acid for many diverse organisms including diatoms<sup>8</sup> and fungi (trichophyton).<sup>9</sup> Sulfanilamide also completely inhibits the growth of normal and mutant strains of *Neurospora*, and in both cases this inhibition is completely overcome by an excess of *p*-aminobenzoic acid. Figure 4 shows that growth on solid medium is completely inhibited by a sulfanilamide concentration of 16 mg. per 15 ml. In these tests enough *p*-aminobenzoic acid was supplied to the mutant to permit normal growth in the absence of sulfanilamide. The quantitative effect of sulfanilamide was the same on the normal and the mutant strains, demonstrating that the utilization of *p*-aminobenzoic acid rather than its synthesis is blocked by sulfanilamide, and that the amount of *p*-aminobenzoic acid synthesized by the normal strain is not appreciably greater than that required by the mutant for normal growth. Similar results were obtained by measurements of dry weight increase. The effectiveness of added *p*-aminobenzoic acid in overcoming the inhibition by 8 mg. sulfanilamide per 25 ml. in liquid medium was also investigated. The vitamin had the same anti-sulfanilamide activity for the normal and mutant strains, but its activity under these conditions was only about  $1/100$  of its vitamin activity for the mutant strain in the absence of sulfanilamide.

As a working hypothesis, a single gene may be considered to be concerned with the primary control of a single specific chemical reaction.<sup>10</sup> If this premise is accepted, the biosynthesis of *p*-aminobenzoic acid in the mutant strain, which differs from normal by a single gene, should be blocked in only one specific step in the entire complex of reactions. An attempt was made to trace the course of *p*-aminobenzoic acid synthesis and to determine the step that is blocked by the mutant gene. This involved testing a number of substances for their ability to replace *p*-aminobenzoic acid in the growth of the mutant strain.<sup>11</sup> The results obtained by the dry weight and growth rate methods as summarized in table 2 were reasonably consistent with each other. *Ortho*- and *m*-aminobenzoic acids were inactive, as were all the compounds tested which did not have an aromatic N. With the amino group replaced by Cl or OH as in *p*-Cl or *p*-OH-benzoic acids, or as in tyrosine, no activity could be detected. These facts indicate that an aromatic substituted N is essential for activity and that if a second substituent group is present it must be in para position to the N. These results and the inactivity of tyrosine, which is presumably synthesized by the mutant strain, suggest the possibility that the reaction blocked by the

mutant gene may be concerned with the introduction of an amino N into the benzene ring.

If this is the reaction blocked in the aminobenzoicless mutant, normal strains should be able to introduce an amino N into a benzene ring. It is known that normal strains synthesize *p*-aminobenzoic acid or some other active substance, and practically all of the excess of this is contained in the culture medium. It is therefore possible to test such strains for their ability to introduce an aromatic amino group by growing them in the presence of compounds containing a nitrogen-free benzene ring and determining whether the *p*-aminobenzoic acid content of the medium is thereby increased. When grown for 3 days in the absence of such compounds, *N. crassa* produced about 0.1 gamma of *p*-aminobenzoic acid per 25 ml. of medium, indicating that the synthesis of this growth factor is normally not greatly in excess of its needs. This is likewise indicated by the sulfanilamide inhibition data. In the presence of benzoic acid, *p*-hydroxybenzoic acid, or tyrosine, in concentrations up to 1 mg. per 25 ml., the amount of *p*-aminobenzoic acid produced was not detectably increased. In fact, benzoic acid appreciably decreased the amount of *p*-aminobenzoic acid recovered. The conclusion that these compounds are not converted into *p*-aminobenzoic acid is confirmed by the fact that they do not have the anti-sulfanilamide activity to be expected with an increased yield of *p*-aminobenzoic acid.

The failure to detect any conversion of the tested nitrogen-free compounds to *p*-aminobenzoic acid, tends to discredit the suggestion that the mutant gene blocks the introduction of an aromatic amino group. It indicates, rather, that the normal synthesis of *p*-aminobenzoic acid does not involve the introduction of an aromatic N or amino group into a pre-formed benzene ring. It should be pointed out, however, that the interpretation of all such experiments may be complicated by factors such as penetration into the cells, secondary chemical reactions or the inability to add the possible intermediates in the biological or "active" condition.

A normal precursor of *p*-aminobenzoic acid which is located in the synthetic sequence leading up to the break due to the mutant gene should theoretically be completely inactive for the mutant strain. Similarly, a precursor coming after this break should be comparable in activity to *p*-aminobenzoic acid itself. If all the effective substances listed in table 2 owe their activity to their conversion to *p*-aminobenzoic acid, *Neurospora* must be able to reduce a nitro group,<sup>12</sup> introduce a carboxyl group para to an amino group, oxidize a methyl or aldehyde group and de-acetylate or de-methylate an amino group. The activities of the effective compounds are so much lower than that of *p*-aminobenzoic acid itself, however, that it seems doubtful that any is its normal precursor, or that any of these reactions is involved in the normal synthesis of *p*-aminobenzoic acid.

These compounds may be somewhat active *per se*, or may owe their activity to a conversion to a *p*-aminobenzoic acid by a relatively inefficient mechanism not concerned in the normal biosynthesis of *p*-aminobenzoic acid. This accessory conversion mechanism should be equally efficient in the normal and mutant strains. In either case the compounds should have the same anti-sulfanilamide activity for both strains if they are not normal precursors of *p*-aminobenzoic acid. Of all the active compounds listed (table 2), only *p*-nitrobenzoic acid and aniline had any anti-sulfanilamide activity,<sup>13</sup> and no quantitative difference in their effect on the normal and mutant strains was observed.

It may be concluded that the active compounds listed in table 2 are probably not concerned in the normal synthesis of *p*-aminobenzoic acid. In this biosynthesis the N or amino group is probably introduced before the formation of the benzene ring.

*Summary.*—An x-ray induced mutant strain of *Neurospora crassa* has been obtained which requires *p*-aminobenzoic acid for growth. Its growth is a function of the amount of *p*-aminobenzoic acid supplied, and it is indistinguishable from normal when adequate amounts of *p*-aminobenzoic acid are available.

The mutant differs from normal by a single gene, which must therefore control an essential step in the synthesis of *p*-aminobenzoic acid, and which is presumably primarily concerned only with the synthesis of *p*-aminobenzoic acid.

Sulfanilamide inhibits the growth of both the normal and mutant strains, and in both cases the inhibition is overcome by an excess of *p*-aminobenzoic acid.

A number of substances related to *p*-aminobenzoic acid are able to replace it, but their activities are much less than that of *p*-aminobenzoic acid itself.

The addition of benzoic or *p*-OH-benzoic acids or tyrosine did not increase the amount of *p*-aminobenzoic acid produced by the normal strain.

It is concluded that none of the compounds tested is concerned with the normal synthesis of *p*-aminobenzoic acid, and that this biosynthesis probably does not involve the introduction of an amino group into a preformed benzene ring.

\* Work supported by grants from the Rockefeller Foundation and from the Penrose Fund of the American Philosophical Society. The authors are indebted to Dr. Russell Perry Hager and Miss Caryl L. Parker for assistance.

<sup>1</sup> See Fildes, P., *Lancet*, 238, 955 (1940); Rubbo, S. D., and Gillespie, J. M., *Nature*, 146, 838 (1940); and Lampen, J. O., and Peterson, W. H., *Jour. Am. Chem. Soc.*, 63, 2283 (1941).

<sup>2</sup> Ansbacher, S., *Science*, 93, 164 (1941); and Sure, B., *Ibid.*, 94, 167 (1941).

<sup>3</sup> Beadle, G. W., and Tatum, E. L., *Proc. Nat. Acad. Sci.*, 27, 499 (1941).

<sup>4</sup> S. M. A. Corporation, Chagrin Falls, Ohio. Biotin concentrate 1000.

<sup>6</sup> Dodge originally used "A" and "a" to designate the two sex types in *Neurospora crassa*. Lindegren later substituted the symbols "+" and "-" for these. Because of the desirability of using the symbol + for the normal allele of a gene and also to designate a positive experimental result, we prefer Dodge's original designations.

<sup>8</sup> Neither conversion of pimelic acid to *p*-aminobenzoic acid by the normal strain nor anti-sulfanilamide activity of pimelic acid could be detected.

<sup>7</sup> See Woods, D., and Fildes, P., *Chem. Ind.*, 59, 133 (1940); Woods, D., *Brit. Jour. Exp. Path.*, 21, 74 (1940); and Landy, M., and Wyeno, J., *Proc. Soc. Exp. Biol. Med.*, 46, 59 (1941).

<sup>8</sup> Wiedling, S., *Science*, 94, 389 (1941).

<sup>9</sup> Dimond, N. S., *Ibid.*, 94, 420 (1941).

<sup>10</sup> See Beadle, G. W., and Tatum, E. L., *Amer. Nat.*, 75, 107 (1941), and loc. cit., footnote 3.

<sup>11</sup> All the compounds tested were carefully purified to remove any possible traces of *p*-aminobenzoic acid. The aniline was prepared by hydrolysis of recrystallized acetanilide and subsequent distillation.

<sup>12</sup> A bacteriostatic action of *p*-nitrobenzoic acid has been reported. See King, J. T., and Henschel, A. F., *Proc. Soc. Exp. Biol. Med.*, 47, 400 (1941).

<sup>13</sup> The other active compounds were toxic in concentrations theoretically high enough to overcome the sulfanilamide inhibition.

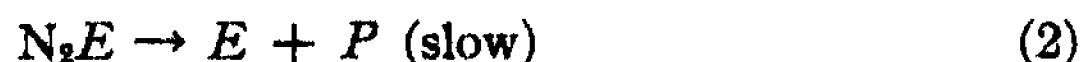
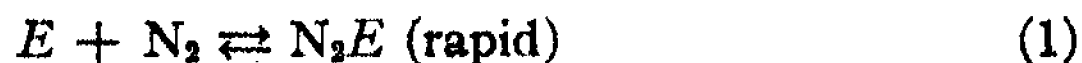
## THE DISSOCIATION CONSTANT IN NITROGEN FIXATION BY AZOTOBACTER

BY P. W. WILSON, R. H. BURRIS AND C. J. LIND

DEPARTMENT OF AGRICULTURAL BACTERIOLOGY, UNIVERSITY OF WISCONSIN

Communicated May 8, 1942

The dissociation constant of nitrogen fixation by *Azotobacter* was first estimated by Lineweaver, Burk and Deming<sup>1</sup> from their data on the influence of the *p*N<sub>2</sub> on rate of fixation by this organism. These workers measured fixation in the Warburg respirometer by observing the increase in the rate of oxygen uptake with time. The initial steps of nitrogen fixation by this organism were formulated as:



where *E* = Enzyme (nitrogenase) concerned with first step of fixation,

*P* = Products (increase in *Azotobacter* cells).

Using the method of Lineweaver and Burk,<sup>2</sup> they calculated a Michaelis constant (*K<sub>N</sub>*) for reaction 1 of 0.21 atm. In their experiments the *p*O<sub>2</sub> was kept constant at 0.2 atm. and hydrogen added to bring the total pressure to one atmosphere whenever the *p*N<sub>2</sub> was less than 0.8 atm. Since



Wilson and his collaborators<sup>1, 4</sup> subsequently demonstrated that molecular hydrogen competitively inhibits both symbiotic and asymbiotic nitrogen fixation, this estimate of the  $K_{N_2}$  must be discarded.<sup>5, 6</sup>

Wyss, *et al.*,<sup>6</sup> have recently calculated a preliminary value for  $K_{N_2}$  based on experiments in which hydrogen was not used. The major purpose of their experiments, however, was to compare the nitrogen fixing system in *Azotobacter* with that in inoculated red clover plants. Consequently, they employed the same technique for preparation and handling of the gas mixtures as had been used previously in tests with the symbiotic system. The  $pN_2$  in the mixtures was probably accurate to no more than 0.01 atm. even at the lower partial pressures of this gas. Although great precision was not claimed for the estimate, Wyss and his associates believed the  $K_{N_2}$  value was near 0.01 to 0.02 atm. Recently we have checked this estimate through use of a number of methods, several of which have not been previously described.

*Fixation at  $pN_2$  0.2 and 0.78 Atm.*—During the symposium on respiratory enzymes held at the University of Wisconsin in September, 1941, the senior author discussed with Dr. Dean Burk the question of the probable value for the Michaelis constant of nitrogen fixation. In the course of our discussions and subsequent correspondence he pointed out that our published data indicated that the rate of fixation (as measured by the monomolecular velocity constant,  $k$ ) at a  $pN_2$  of 0.2 atm. did not differ significantly from that in air. He suggested that a difference should be detectable unless the Michaelis constant was rather small since the familiar Michaelis-Menten equation states

$$k/k_{\max.} = pN_2/(pN_2 + K_{N_2}). \quad (3)$$

Substitution of possible values for  $K_{N_2}$  into this equation leads to the following:

$K_{N_2}$	0.2 ATM.	$k/k_{\max.}$	$k$ AT 0.2 ATM.	
			AIR	$k$ IN AIR
0.01	95.3%		98.7%	96.5%
0.02	90.9%		97.5%	93.2%
0.03	87.0%		96.3%	90.3%

This observation suggested to us a method for determining the probable limits of  $K_{N_2}$  provided that reliable estimates of the error in the  $k$  values were available. Our values of the  $k$ 's had been calculated from the slopes of straight lines by the usual statistical procedures; hence their errors were known. For the macronitrogen experiments each value of  $k$  was based on six observations. From 25 such lines the average error in the estimate of  $k$  was calculated as  $5.72 \pm 0.41$  per cent. Even though the  $K_{N_2}$  were as much as 0.03 atm., a *significant* difference would not always be obtained

in an *individual* experiment. If, however, a large number of experiments were combined, a real difference should be evident, and its magnitude should indicate the probable size of  $K_{N_2}$ . Data from 18 experiments were available. The mean value for  $k$  in air was 0.0715, in a  $pN_2$  of 0.2 atm., 0.0688; the mean difference was  $0.0027 \pm 0.0009$ . The  $k$  values for the  $pN_2$  of 0.2 atm. ranged from 87 to 108 per cent of the corresponding value in air with an average of  $96.7 \pm 1.23$ . This value suggests a Michaelis constant of 0.01 but is consistent with one as high as 0.02 atm.

The microrespirometer method offers the advantage that a large number of determinations can be readily made for estimation of  $k$  with consequent reduction in its error. Whereas it is feasible to base the  $k$  values on only 6 determinations in the macronitrogen experiments, duplicate flasks on which 10 readings are taken over a period of 5 hours are easily handled by the microrespirometer technique. In the first instance the  $k$  value has 4 degrees of freedom, in the second, 16; a substantial reduction in the error of estimate would be expected. The average error in 22 such estimates of  $k$  from microrespiration experiments was  $3.71 \pm 0.24$  per cent. Our previous studies included just three trials with a  $pN_2$  of 0.2 atm., and the  $k$  values from these had only 9–12 d. f. None was significantly different from the air control; the  $k$  values from a large number of trials with a  $pN_2$  of 0.3 atm. also were identical within experimental error with those in air. Three new experiments were made in which triplicate flasks of each treatment were taken and 10 readings made on each. The mean values of the velocity

TABLE 1

EFFECT OF  $pN_2$  ON REACTION VELOCITY CONSTANT IN MICRORESPIRATION EXPERIMENTS

$p'N_2$ IN ATM.	$k$	$k'$	$k - k'$	$K_{N_2}$
0.214	$0.245 \pm 0.0080$	$0.219 \pm 0.0072$	$0.026 \pm 0.011$	$0.037 \pm 0.017$
0.019		$0.120 \pm 0.0055$	$0.125 \pm 0.009$	$0.021 \pm 0.0023$
0.214	$0.233 \pm 0.0099$	$0.219 \pm 0.0076$	$0.014 \pm 0.012$	$0.020 \pm 0.018$
0.019		$0.095 \pm 0.0054$	$0.138 \pm 0.011$	$0.029 \pm 0.0037$
0.214	$0.242 \pm 0.0049$	$0.215 \pm 0.0082$	$0.027 \pm 0.009$	$0.039 \pm 0.014$
0.019		$0.090 \pm 0.0099$	$0.152 \pm 0.011$	$0.034 \pm 0.0065$

$k$ , velocity constant in air;  $k'$  in indicated  $p'N_2$ .

constants were thus based on 24 d. f. so that a difference of about 8% in two means would be significant. The results in table 1 show that in two experiments a significant difference was evident; in the other, the difference was in the right direction but did not exceed experimental error.

*Calculation from Formula.*—If  $k$  represents the rate constant for fixation in air and  $k'$  that at some other partial pressure of nitrogen,  $p'N_2$ , the following relation can be derived from equation (3):

$$K_{N_2} = \frac{k - k'}{k'/p' - k/p} \quad (4)$$

TABLE 2

THE  $pN_2$  FUNCTION OF NITROGEN FIXATION BY *Azotobacter* IN MICRORESPIRATION  
EXPERIMENTS†

$pN_2$ IN ATM.	EXPERIMENT 1	EXPERIMENT 2	EXPERIMENT 3
0.01	0.139 } 0.113 } 0.126	0.107 } 0.098 } 0.103	0.146 } 0.139 } 0.142
0.015	0.131 } 0.130 } 0.130		0.155 } 0.156 } 0.155
0.025	0.173 } 0.171 } 0.172	0.173 } 0.167 } 0.170	0.202 } 0.209 } 0.205
0.05	0.192 } 0.181 } 0.186	0.184 } 0.166 } 0.175	0.230 } 0.223 } 0.227
0.10	0.232 } 0.212 } 0.222	0.197 } 0.185 } 0.191	0.246 } 0.242 } 0.244
0.20	0.234 } 0.230 } 0.232	0.216 } 0.202 } 0.209	0.253 } 0.259 } 0.256
0.78		0.204 } 0.235 } 0.220	
Significant difference*	0.017	0.015	0.013

\* Difference between means necessary for significance.

† Values in the table are reaction velocity constants—the  $k$  values.

Values of  $K_{N_2}$  given in table 1 were calculated from this formula. In addition to the experiments already described, parallel flasks were included in which the  $pN_2$  was 0.019 atm. This atmosphere was prepared by combining 0.8 atm. of tank helium with 0.2 atm. of pure oxygen. According to the manufacturer, the tank helium contains about 1.8%  $N_2$ ; this value had been used in our previous experiments. For these experiments, however, more precise control of the  $pN_2$  was believed essential so the nitrogen content in the particular tank of helium used was determined by an application of the isotope dilution principle.<sup>7</sup> Known quantities of tank helium and nitrogen gas containing an excess of the  $N^{15}$  isotope were mixed, and the isotope dilution effected by the normal nitrogen in the tank helium was measured by analysis in the mass spectrometer. Using two widely different mixtures, we obtained values of 2.29 and 2.46%  $N_2$  in the helium. The values of the  $K_{N_2}$  calculated from these data were definitely higher than those usually obtained. It should be observed, however, that, except for one case, the high values have a very large error\* so that they are not inconsistent with a  $K_{N_2}$  of about 0.02 atm.

As has been mentioned, no special precautions were taken in our previous experiments to control [within narrow limits] the  $pN_2$  in the atmospheres. The effect of this source of error would be increased as the  $pN_2$  was decreased. When the latter exceeded 0.05 atm., a variation 0.01 would not markedly alter the estimate of  $K_{N_2}$ . Accordingly, values of this con-

stant were calculated by equation (4) from all trials in which a  $pN_2$  of 0.05 to 0.2 atm. had been used. The results were:

TYPE OF EXPERIMENT	NUMBER	$K_{N_2}$
Microrespirometer	16*	$0.029 \pm 0.0020$ atm.
Macronitrogen (Wyss)	12	$0.013 \pm 0.0014$
Macronitrogen (Lind)	16	$0.020 \pm 0.0025$

\* Includes the 6 trials in table 1.

*The  $pN_2$  Function.*—As check on these values we carried out new experiments in which elaborate precautions were used to control the  $pN_2$  in the

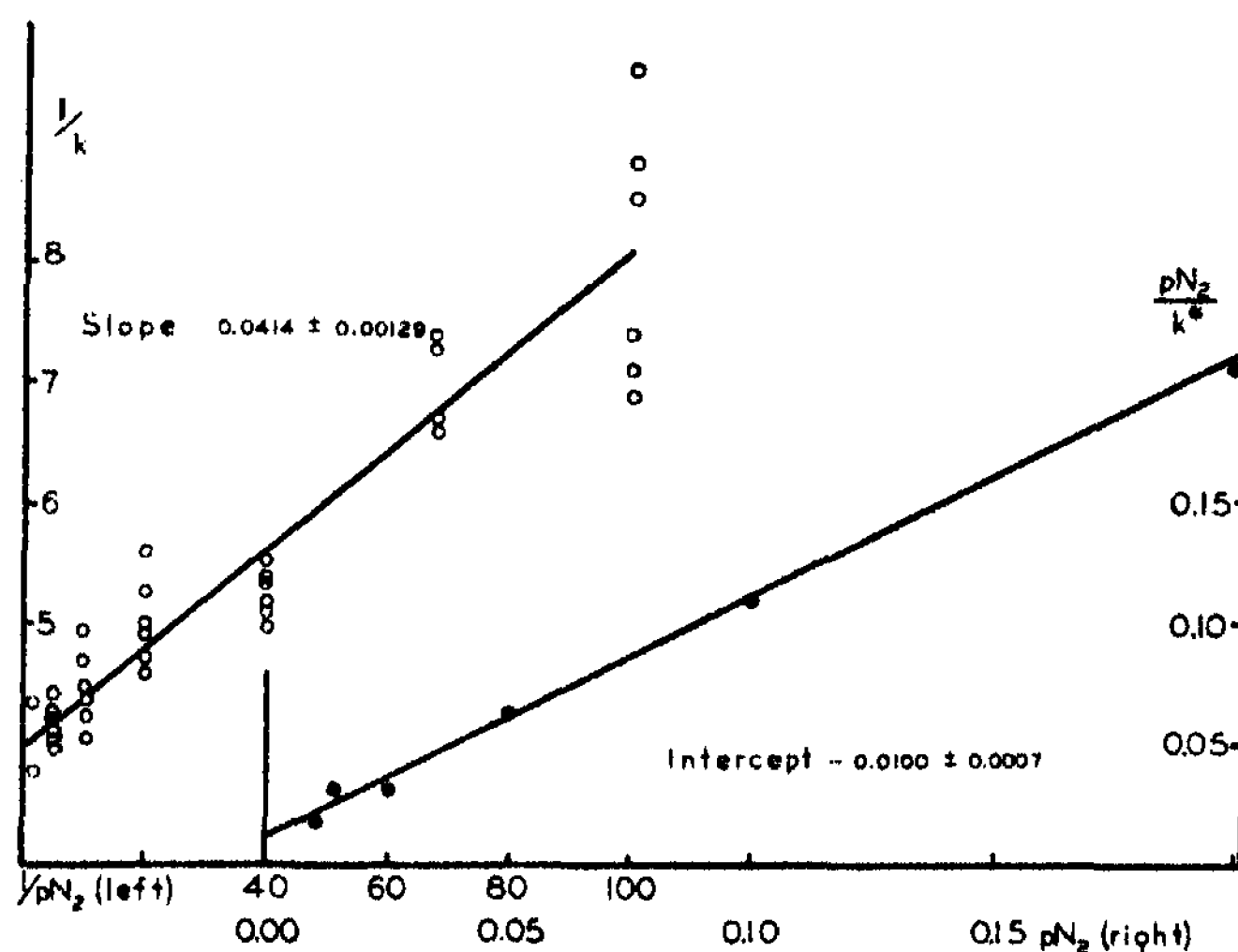


FIGURE 1

Estimation of  $K_{N_2}$  by methods based on equations (5) and (7).

(Data of microrespiration experiments. Since the scatter about the line was too small to allow plotting of individual points, only the mean values are shown in line at right; all values are plotted for line at left. The apparent worse fit of these points to the "best" line arises primarily from difference in scale.)

atmospheres. A source of pure helium was used (100% according to manufacturer), and the required atmospheres prepared over mercury with a Toepler pump using pure oxygen generated from potassium chlorate. The gases were transferred to containers which had been evacuated to  $10^{-3}$  mm. Hg. The Warburg flasks were evacuated to 25 mm. Hg residual pressure and pure oxygen added; this was repeated three times, then the desired gas mixture added after the fourth evacuation. Water, autoclaved just before use, served as the displacing fluid. In spite of every care, it is believed that the nitrogen content in the atmospheres supplied the azoto-

bacter was slightly higher than calculated since helium-oxygen controls showed a small but definite increase in respiration with time. There was no increase in the respiration of the hydrogen-oxygen controls in which fixation would not occur even though a small quantity of nitrogen accidentally did get into the mixture. The estimates of  $K_{N_2}$ , based on the data in table 2, therefore, are probably low and should be regarded as minimum values.

One estimate of the  $K_{N_2}$  was made from these data by the method of Lineweaver and Burk<sup>2</sup> based on the equation:

$$1/k = (K_{N_2}/k_{\max.})(1/pN_2) + 1/k_{\max.} \quad (5)$$

The value for  $1/k_{\max.}$  (intercept) was determined for each experiment, then the individual values adjusted to a common intercept of 4.0. The combined data were fitted to a line with this intercept, and the  $K_{N_2}$  estimated by dividing the slope by 4.0; test of the line showed no evidence of departure from linearity (Fig. 1). A value of  $0.0104 \pm 0.00032$  atm. was obtained.

A second method for estimating  $K_{N_2}$  from such data was suggested by Burk (personal communication). If equation (5) is multiplied by  $(pN_2)$  it becomes:

$$pN_2/k = pN_2/k_{\max.} + K_{N_2}/k_{\max.} \quad (6)$$

If  $k_{\max.}$  is taken as unity, this equation reduces to:

$$pN_2/k^* = pN_2 + K_{N_2} \quad (7)$$

in which  $k^*$  is the velocity constant relative to  $k_{\max.} = 1.00$ . In each experiment  $k_{\max.}$  was determined by the method based on equation (5), then *relative*  $k$  values calculated from the observed  $k$ 's by dividing each by the proper  $k_{\max.}$  Values of  $pN_2/k^*$  were plotted against the  $pN_2$  and a straight line with unit slope fitted to the points. The intercept,  $0.0100 \pm 0.0007$ , determines  $K_{N_2}$  (Fig. 1).

*Inhibition Experiments.*—Burk<sup>3</sup> observed that the slope/intercept from the lines based on the experiments in which hydrogen was the replacing gas did not determine  $K_{N_2}$  as originally supposed but a more complex relationship:

$$K_{H_2}/K_{N_2} = 5(1 + K_{H_2}) \text{ or} \quad (8)$$

$$= 5/(1 - 5K_{N_2}). \quad (9)$$

Since  $K_{N_2}$  is likely between 0.01 and 0.02 atm., the value of  $K_{H_2}/K_{N_2}$  is close to 5.4; also moderate change in  $K_{N_2}$  would not markedly alter the value of the ratio. From the equation of competitive inhibition by hydrogen,<sup>4</sup> the following relationship is derived:

$$K_{N_2} = \frac{pH_2/(K_{H_2}/K_{N_2}) - pN_2(k/k_i - 1)}{(k/k_i - 1)} \quad (10)$$

in which at a given  $pN_2$ ,  $k$  and  $k_i$  are the rates in the absence and presence of hydrogen, respectively. Ten experiments were available which allowed test of this equation. As the error of the estimate was rather large, it was not surprising that the calculated values of  $K_{N_2}$  fell on either side of zero. Three experiments were discarded since the estimates of  $K_{N_2}$  were obviously too large—order of  $\approx 0.10$  atm. The mean of the remaining seven was  $0.014 \pm 0.0066$  atm.

*Discussion.*—In evaluating the estimates of  $K_{N_2}$  from the several types of experiments two sources of error must be distinguished—experimental errors in the determinations and absolute errors caused by the particular technique used. The first of these has been taken care of by the statistical treatment of the data. Six to thirty determinations were made in each trial, and several experiments were combined for the final estimate. As a result the experimental error was quite small except that based on the inhibitor experiments. Sources of possible absolute error are more difficult to recognize and eliminate. One has been discussed in the text: the variation in the  $pN_2$ . A second is that the velocity constant of fixation is indirectly measured in the microrespiration experiments. Although nitrogen fixed and increase in oxygen uptake undoubtedly roughly parallel each other, assurance is lacking that the relationship remains linear as the experimental conditions change, e.g., as the  $pN_2$  is altered. In the macro-nitrogen experiments secondary effects arising from aging of the culture may influence the results. The fact that a constant rate is obtained when the logarithm of the nitrogen fixed is plotted against time suggests, however, that in the 24 to 36 hours required for these trials, such effects were not marked.

Consideration of all estimates of  $K_{N_2}$  suggests  $0.02 \pm 0.005$  atm. as the most probable value. The estimate from the most precise experiments was 0.01 atm., but there is reason to believe that this is a minimum figure. The highest estimate was 0.029 atm. Since this value is based on experiments in some of which no special care was taken to control the  $pN_2$  (variation of at least 0.01 atm.), it probably represents a maximum. The remaining three estimates are consistent with 0.02 atm. We are inclined to the view that the lower limit of the suggested value, i.e., 0.01 to 0.015 atm., is more probable than is the indicated upper limit 0.025 to 0.03 atm., since the very high values were observed only in microrespiration experiments in which the measurement of fixation is indirect. Some variation would be expected because of the different methods used. The relatively close agreement among the figures from different types of experiments and methods of estimation lends confidence to the results and to the mechanism

hypothesis on which they are based. If 0.02 atm. is taken as most likely for the  $K_N$ , then  $K_H$  becomes  $0.11 \pm 0.028$  atm., a value in good agreement with that previously made by Wyss, *et al.*<sup>6</sup>

*Summary.*—Four methods for estimating the dissociation (Michaelis) constant of an enzyme-substrate complex are illustrated with data from experiments on nitrogen fixation by *Azotobacter vinelandii*. The values found for this reaction ranged from 0.01 to 0.029 atm.; based on present results,  $0.02 \pm 0.005$  atm. is suggested as the most probable estimate. The corresponding value for the  $K_H$ , i.e., for the dissociation constant of the enzyme-inhibitor complex, is  $0.11 \pm 0.028$  atm.

The authors express their appreciation to Dr. Dean Burk for his helpful suggestions and especially for permission to use the method based on equation (7). We are also indebted to Dr. Churchill Eisenhart, Station Statistician, for advice on the several statistical procedures used including the method for estimating the correlation between  $(k - k')$  and  $(k'/p' - k/p)$ , i.e.,  $a$  and  $b$ . The mass spectrometric analyses were made by F. J. Eppling through the courtesy of Professor H. B. Wahlin of the Physics department.

\* Note that  $K_N$  is the ratio of two quantities,  $a/b$ , each subject to errors which are correlated as follows:  $r = -(p'\sigma_k^2 + p\sigma_{k'}^2)/(pp'\sigma_a\sigma_b)$ . The estimate of  $K_N = a/b(1 - r\sigma_a\sigma_b/ab + \sigma_b^2/b^2)$ , as well as its error, was determined by the method described by Yule<sup>8</sup> for determination of mean and standard deviation of an index.

<sup>1</sup> Lineweaver, H., Burk, D., and Deming, W. E., *Jour. Amer. Chem. Soc.*, **56**, 225-230 (1934).

<sup>2</sup> Lineweaver, H., and Burk, D., *Ibid.*, **56**, 658-666 (1934).

<sup>3</sup> Wyss, O., and Wilson, P. W., *Proc. Nat. Acad. Sci. (U. S.)*, **27**, 162-168 (1941).

<sup>4</sup> Wilson, P. W., *The Biochemistry of Symbiotic Nitrogen Fixation*, University of Wisconsin Press, Madison, Wis., 1940, 302 pp.

<sup>5</sup> Burk, D., and Burris, R. H., *Ann. Rev. Biochem.*, **10**, 587-618 (1941).

<sup>6</sup> Wyss, O., Lind, C. J., Wilson, J. B., and Wilson, P. W., *Biochem. Jour.*, **35**, 845-854 (1941).

<sup>7</sup> Rittenberg, D., and Foster, G. L., *Jour. Biol. Chem.*, **133**, 737-744 (1940).

<sup>8</sup> Yule, G. U., *An Introduction to the Theory of Statistics*, 8th ed., London, 1927, Chapter 11.



## SOME DEDUCTIONS FROM FROBENIUS'S THEOREM

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated May 16, 1942

In 1895 G. Frobenius announced the theorem that if  $m$  is a factor of the order of a finite group  $G$  then the number of the operators of  $G$ , including the identity, whose orders divide  $m$  is a multiple of  $m$ . A necessary and sufficient condition that  $G$  is cyclic is that this number is equal to  $m$  for every divisor of the order of  $G$ . This results directly from the facts that every subgroup of a cyclic group is cyclic and that a cyclic group contains one and only one subgroup whose order is an arbitrary divisor of the order of the group. Whenever  $G$  is non-cyclic its order is divisible by at least one positive integer such that the number of the operators of  $G$  whose orders divide this integer exceeds it. This number is clearly not equal to unity or to the order of  $G$ . In the particular case when  $G$  is a dihedral group it is easy to see that whenever a factor of the order of  $G$  is odd then  $G$  contains exactly as many operators whose orders divide this factor as there are units in the factor but whenever such a factor is even and less than the order of  $G$  then the number of the operators of  $G$  whose orders divide this factor exceeds it.

To illustrate this theorem of Frobenius it may be desirable to prove it separately for the case when  $G$  is any abelian group. This may also throw additional light on the characteristic subgroups of this important class of groups. It is obvious that every abelian group contains at least one subgroup whose order  $m$  is an arbitrary divisor of the order of the group. Since the order of a subgroup is divisible by the order of each of its operators it results that when  $G$  is abelian and  $m$  is a divisor of the order of  $G$  then  $G$  contains at least  $m$  operators whose orders divide  $m$ . A necessary and sufficient condition that  $G$  contains no other operator whose order divides  $m$  is that this subgroup of order  $m$  is composed of all the operators of  $G$  whose orders divide  $m$ . In general, all such operators constitute a special characteristic subgroup of  $G$  whose order is a multiple of  $m$  because the order of a group is a multiple of the order of each of its subgroups. This proves the theorem of Frobenius for abelian groups.

It is not difficult to determine the number of the characteristic subgroups of the abelian group  $G$  which constitute the given category. In the special case when the order of  $G$  is a power of a prime, say  $p^\alpha$ , the number of these subgroups is clearly  $\beta + 1$  if  $p^\beta$  is an operator of highest order contained in  $G$  and the identity and the group itself are included among the characteristic subgroups. In the general case the order of  $G$  may be assumed to be  $p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$ ,  $p_1, p_2, \dots, p_k$  being distinct prime numbers.



If we let  $p_1^{\beta_1}, p_2^{\beta_2}, \dots, p_k^{\beta_k}$  be the orders of the largest operators in the Sylow subgroups of  $G$  then the number of the given characteristic subgroups of  $G$ , including the group itself and the identity, is  $(\beta_1 + 1)(\beta_2 + 1) \dots (\beta_k + 1)$ . Hence *this is the number of the factors of the order of  $G$  which have the property that they are equal to the number of the operators of  $G$  whose orders divide them respectively, whenever  $G$  is abelian.* For all other factors of the order of  $G$  there are more operators of  $G$  whose orders divide these factors than there are units in these respective factors.

Suppose now that  $G$  contains a cyclic Sylow subgroup of order  $p^m$ . If such a  $G$  contains a subgroup of index  $p$  this subgroup is composed of all the operators of  $G$  whose orders are not divisible by  $p^m$  since it is then generated by these operators and any one of its operators of order  $p^m$ . In every case the number of the operators of  $G$  whose orders are not divisible by  $p^m$  cannot exceed the order of  $G$  multiplied by  $(p - 1)/p$  according to the theorem noted at the opening of this article since this number must be a multiple of the order of  $G$  divided by  $p$  and cannot be equal to the order of  $G$ . In the particular case when  $p = 2$  it is obviously equal to this number and it is then easy to see that these operators constitute a subgroup of index 2 under  $G$  since they are composed of the positive permutations of  $G$  when  $G$  is represented as a regular permutation group. These positive permutations then constitute an invariant subgroup of index 2 under  $G$ . In the special case when the order of  $G$  is twice an odd number this reduces to the well-known theorem that the operators of odd order in such a group constitute an invariant subgroup under the group.

Every transitive permutation group of degree  $n$  is known to have an order which is a multiple of  $n$  since it contains a subgroup of index  $n$  composed of all its permutations which omit a given letter. When the group is non-regular its order is therefore always divisible by a number which exceeds unity and is less than the order of the group. The number of its permutations whose orders divide this number must therefore be a multiple of the order of this subgroup composed of all its permutations which omit a given letter according to the theorem under consideration. In the special cases when  $G$  is either symmetric or alternating this theorem can readily be proved directly. In the former case when  $n$  is composite and exceeds 4 it is easily seen that all the permutations of the group have orders which divide  $(n - 1)!$  since every substitution of degree  $n$  contained therein has this property. When  $n$  is a prime number  $p$  then the number of the permutations of degree  $p$  is obviously  $(p - 1)!$  and their orders do not divide  $(n - 1)!$

Hence it results that in the symmetric group of degree  $n$  all the operators have orders which divide the order of its subgroup composed of all its permutations which omit a given letter whenever  $n$  is composite and exceeds 4. When  $n$  is prime then all of these operators except  $(n - 1)!$  have this

property. This is also clearly the case when  $n = 4$ . When  $G$  is the alternating group of degree  $n$  and  $n$  is a composite number greater than 4 it follows for the same reason as in the case of the symmetric group that the orders of all its operators divide the order of its subgroup of index  $n$ , but when  $n$  is a prime number there are twice as many operators in the alternating group whose orders do not divide the order of its subgroup of index  $n$  as there are operators in this subgroup since the number of these operators is the same as in the case of the corresponding symmetric group but the order of the subgroup is only half as large.

If in an odd prime power group the number of the operators whose orders divide one of the factors of the order of the group which exceeds unity but is less than the order of the group is exactly equal to this factor then this is also true as regards the number of operators whose orders divide any other such factor. This results directly from the known fact that such a group is necessarily cyclic when it satisfies the given condition because it contains one and only one proper subgroup of every order which exceeds unity and is less than the order of the group.<sup>1</sup> Similar remarks apply to the groups of order  $2^m$  except that in this case the given factor must exceed 2 since the dicyclic group of order  $2^m$  contains only one subgroup of order 2. It is obvious that a necessary and sufficient condition that a group of order  $p^m$ ,  $p$  being a prime number, is cyclic is that it contains operators whose orders do not divide  $p^{m-1}$ .

A necessary and sufficient condition that the group  $G$  contains an invariant Sylow subgroup of order  $p^m$  is that it contains exactly  $p^m$  operators whose orders are powers of  $p$ . If it contains more than  $p^m$  such operators it contains at least  $p^m + 1$  such operators since it must then contain at least  $p + 1$  Sylow subgroups of order  $p^m$  and if it contains more than  $p + 1$  such subgroups it also contains more than  $p^m + 1$  operators whose orders are powers of  $p$ . According to the given theorem of Frobenius it must then contain  $p^m + 1 + kp^m$  operators whose orders are powers of  $p$ . The number of these operators diminished by unity is obviously divisible by  $p - 1$  and hence  $k$  is divisible by  $p - 1$ . That is, the number of the operators of  $G$  whose order is a power of  $p$  is  $(1 + l)p^m + 1 - lp^m$ , where  $k = l(p - 1)$ . When  $l = 1$  this reduces to  $2p^m + 1 - p^m$ , which is therefore the smallest number of operators in  $G$  whose orders are a power of  $p$  whenever this number exceeds  $p^m + 1$ .

From the theorem noted at the opening of this article it results directly that the number of the operators of a group whose orders do not divide a factor of the order of a given group is either zero or a multiple of this factor. This form of the theorem is sometimes useful when it is employed in determining the total number of the abstract group of a given order. In particular, when  $G$  is a group of order 12 then the number of its operators whose orders divide 4 is either 4 or 8 since not all of its operators can have

orders which divide 4 in view of Sylow's theorem. For the same reason the number of its operators whose orders do not divide 4 is either 8 or 4. Hence a group of order 12 has either 4 or 8 operators whose orders are divisible by 3. If it contains 8 such operators it must involve four subgroups of order 3 and hence it must be the alternating group of order 12 since it transforms these subgroups transitively. If it has only 4 such operators it must involve the cyclic group of order 6 as an invariant subgroup and hence if it is abelian it is either cyclic or the direct product of the cyclic group of order 6 and the group of order 2. If it is non-abelian it is either dihedral or dicyclic. The determination of the five possible groups of order 12 results therefore very easily from the given theorem. This is also true of the fifteen possible groups of order 24.

<sup>1</sup> Miller, Blichfeldt, Dickson, *Finite Groups*, p. 128 (1916).

## ON MONOTHETIC GROUPS

BY PAUL R. HALMOS AND H. SAMELSON

THE INSTITUTE FOR ADVANCED STUDY

Communicated April 17, 1942

§ 1. A topological group  $G$  is called *monothetic* (following van Dantzig<sup>1</sup>) if there exists a cyclic subgroup  $H$  which is dense in  $G$ , i.e., the closure of  $H$  is  $G$ . A generating element of such a cyclic subgroup is called a *generator* of  $G$ .

All groups considered in the sequel are abelian; monothetic groups are evidently abelian.

The elements of finite order of a discrete group  $K$  form a subgroup, the *torsion group*  $T(K)$  of  $K$ .

We call a group  $G$  *separable* if there exists a countable subset which is dense in  $G$ .

We use the theory of character groups. We denote by  $C$  the value group for the characters, the group of real numbers mod 1 with the usual topology; let  $\bar{C}$  be the same group, but with the discrete topology. We denote by  $h$  the natural mapping of  $\bar{C}$  on  $C$ ; it is continuous and an algebraic (but not of course topological) isomorphism.

The character group of a group  $G$  is denoted by  $G^*$ . The *annihilator* of a subset  $H$  of  $G$  is the set of those characters of  $G$  which map every element of  $H$  into the zero element of  $C$ ; it is a (closed) subgroup of  $G^*$ ; we denote it by  $A(H)$ . We recall that the character group of a compact (discrete) group is discrete (compact); for both types of groups we have

the duality theorem, which says  $G^{**} = G$ , and more generally,  $A(A(H)) = H$  for every closed subgroup  $H$ .

§ 2. We restrict our considerations to locally compact groups (as is customary in the theory of abelian groups). It is known that a locally compact monothetic group is either compact or discrete.<sup>2</sup> The discrete case being trivial we consider only the compact case. We prove

**THEOREM I.** *A compact group  $G$  is monothetic if and only if its character group  $G^*$  is isomorphic to a subgroup of  $\bar{C}$ .*

*Proof.* (a) Suppose  $G$  is monothetic; let  $d$  be a generator of  $G$ . Let  $f = f(x)$  ( $x \in G, f(x) \in C$ ) be an arbitrary character of  $G$ ; the mapping  $f \rightarrow h^{-1}(f(d))$  is obviously an isomorphism of  $G^*$  into  $\bar{C}$ , because every character  $f$  is completely determined by its value  $f(d)$  for the generator  $d$  of  $G$ .

(b) Suppose  $G^*$  is a subgroup of  $\bar{C}$ . The mapping  $h$  of  $\bar{C}$  into  $C$  induces a homomorphic mapping of  $G^*$  into  $C$ , i.e., a character of  $G^*$ ; call this character  $d$ . Since  $G = (G^*)^*$ ,  $d$  may be considered as an element of  $G$ . Let  $D$  be the subgroup of  $G$  generated by  $d$ , i.e., the closure of the cyclic group generated by  $d$ . The character  $d$  maps only the zero element of  $G^*$  into the zero of  $C$ ; this means that the annihilator of  $d$ , and hence also that of  $D$ , contains only the zero of  $G^*$ . But this means obviously that  $D$  equals  $G$ . This proves Theorem I.

(c) From the considerations in (a) and (b) it is clear that an element  $d$  of  $G$  is a generator of  $G$  if and only if its annihilator contains only the zero element of  $G^*$ .

(d) Suppose again  $G^*$  is a subgroup of  $\bar{C}$ . It can be shown easily by elementary methods, without using the duality theorem, that the cyclic group generated by the character  $d$  of  $G^*$ , defined in (b), is dense in the character group  $G^{**}$  of  $G^*$ , by direct consideration of the neighborhoods of an arbitrary element of  $G^{**}$ .

§ 3. **THEOREM II.** *A discrete group  $H$  is isomorphic to a subgroup of  $\bar{C}$  if and only if its power (cardinal number) is  $\leq c$  (= the power of the continuum) and its torsion group  $T(H)$  is isomorphic to a subgroup of  $\bar{C}$  and so of  $T(\bar{C})$ .*

Theorem II is a consequence of the following

**LEMMA.** *Let  $H$  be a discrete group of power  $\leq c$ ; every isomorphic mapping  $f_1$  of  $T(H)$  into  $\bar{C}$  can be extended to an isomorphic mapping  $f$  of  $H$  into  $\bar{C}$ .*

*Proof.* The complement of  $T(\bar{C})$  contains  $c$  linearly independent elements (elements of a Hamel basis for the real numbers, reduced mod 1); let  $\lambda_1, \lambda_2, \dots$  be a well ordering of these elements. We well order also the elements of the complement of  $T(H)$ :  $x_1, x_2, \dots$ . We consider the groups  $H_\alpha$  generated by the elements of  $T(H)$  and the  $x_\beta$  with  $\beta < \alpha$ , and construct isomorphic mappings  $f_\alpha$  of  $H_\alpha$  into  $\bar{C}$  such that  $f_\alpha$  is an extension of  $f_\beta$  for  $\beta < \alpha$ . We start with the given  $f_1$ . Suppose  $f_\beta$  is constructed for  $\beta < \alpha$ .

(a) If  $\alpha$  is a limit ordinal then  $H_\alpha$  is the union of the  $H_\beta$  with  $\beta < \alpha$ .

Every  $z \in H_\alpha$  is contained in some  $H_\beta$ ,  $\beta < \alpha$ ; we define  $f_\alpha(z) = f_\beta(z)$ . It is clear that this  $f_\alpha$  has the desired properties.

(b) If  $\alpha$  is not a limit ordinal then  $H_\alpha$  is generated by  $H_{\alpha-1}$  and  $x_{\alpha-1}$ .

(b') If no multiple of  $x_{\alpha-1}$  belongs to  $H_{\alpha-1}$ , then  $H_\alpha$  is the direct sum of  $H_{\alpha-1}$  and the cyclic group generated by  $x_{\alpha-1}$ ; we extend  $f_{\alpha-1}$  by putting  $f_\alpha(x_{\alpha-1}) = \lambda_{\alpha-1}$ . This gives an isomorphic extension, because under our constructions  $\lambda_\gamma$  and the elements of the groups  $f_\beta(H_\beta)$  with  $\beta \leq \gamma$  are always linearly independent.

(b'') If a multiple of  $x_{\alpha-1}$  belongs to  $H_{\alpha-1}$ , let  $n$  be the smallest positive integer for which  $nx_{\alpha-1} \in H_{\alpha-1}$ . Every element  $z$  of  $H_\alpha$  can then be written in a unique manner as  $y + mx_{\alpha-1}$  with  $y \in H_{\alpha-1}$  and  $0 \leq m < n$ . Let  $\lambda$  be an element of  $\bar{C}$  with  $n\lambda = f_{\alpha-1}(nx_{\alpha-1})$ . Put  $f_\alpha(z) = f_{\alpha-1}(y) + m\lambda$ . It is easily verified that  $f_\alpha$  is a homomorphic mapping of  $H_\alpha$ , and that it is an extension of  $f_{\alpha-1}$ . To prove that it is an isomorphism, suppose  $f_\alpha(z) = 0$ . We have then  $nz = 0$ , because  $nz \in H_{\alpha-1}$ ,  $f_{\alpha-1}(nz) = nf_\alpha(z) = 0$ , and  $f_{\alpha-1}$  is isomorphic. So we have  $z \in T(H)$ , and  $f_1(z) = f_\alpha(z) = 0$ ; but  $f_1$  is isomorphic, and so  $z = 0$ .

The desired extension  $f$  of  $f_1$  is now given by  $f(x) = f_1(x)$  for  $x \in T(H)$  and  $f(x_\alpha) = f_{\alpha+1}(x_\alpha)$  for the  $x_\alpha$  which form the complement of  $T(H)$ .

§ 4. We come now to a theorem which is the "dual" of Theorem II.

**THEOREM II\*.** *Let  $G$  be a compact group and let  $G_1$  be its component of the identity;  $G$  is monothetic if and only if it is separable and the totally disconnected factor group  $G/G_1$  is monothetic.*

For the proof we note that the (obviously necessary) separability guarantees that the power condition of Theorem II is fulfilled for  $H = G^*$ . Therefore, by Theorem II  $G^*$  is isomorphic to a subgroup of  $\bar{C}$  if and only if  $T(G^*)$  is. But  $T(G^*)$  is, as is well known, the character group of  $G/G_1$ . Applying now Theorem I to  $G$  and  $G/G_1$  we obtain Theorem II\*.

From this follows immediately the

**COROLLARY.** *Every compact connected separable (abelian) group is monothetic.*

A discrete group is called *locally cyclic* if every subgroup which is generated by a finite number of elements can be generated by a single element. It is easy to see that a group without elements of infinite order is locally cyclic if and only if it is isomorphic to a subgroup of  $T(\bar{C})$ . We may accordingly restate Theorem II\* as

**THEOREM II'.** *A compact group  $G$  is monothetic if and only if it is separable and the torsion group  $T(G^*)$  of its character group is locally cyclic.*

Let  $Z$  be a compact totally disconnected monothetic group. Since the character group of a totally disconnected group has no elements of infinite order it follows from Theorem I that  $Z^*$  is (isomorphic to) a subgroup of  $T(\bar{C})$ . Consequently  $Z^*$  is the direct sum of groups  $Z^*_p$ ,  $p$  running over the prime numbers, where each  $Z^*_p$  is either the zero group or cyclic of order  $p^n$ .

for some  $n$  or isomorphic to the group  $T_p$  of all elements of  $T(\overline{C})$  the order of which is a power of  $p$ . It follows that  $Z$  is the direct sum of groups  $Z_p$ ,  $p$  running over the primes, where  $Z_p$  is either the zero group or cyclic of order  $p^n$  for some  $n$  or the  $p$ -adic group, the  $p$ -adic group being the character group of the group  $T_p$  just mentioned. Conversely, every such direct sum is a compact totally disconnected monothetic group.

§ 5. We prove now a theorem on the Haar measure of the set of generators of a group which is a partial strengthening of the corollary of § 4.

THEOREM III. *The set of generators of a compact connected (abelian) group satisfying the second countability axiom has Haar measure 1.*<sup>3</sup>

*Proof.* The character group  $G^*$  is countable; let  $0, x_1, x_2, \dots$  be its elements. Let  $G_i$  be the set of those elements of  $G$  which are mapped by the character  $x_i$  of  $G$  into the zero of  $C$ . Each  $G_i$  is a closed proper subgroup of  $G$ , and so of measure 0 (because,  $G$  being connected,  $G_i$  has an infinite number of disjoint cosets of equal measure); the union of the  $G_i$  has therefore measure 0 too. Using now remark (c) of §2 we see that the set of generators of  $G$  is identical with the complement of the union of the  $G_i$ , and so of measure 1.

Theorem III is not necessarily true for a group which does not fulfil the second countability axiom. An example is the toral group of dimension  $c$  (the direct sum of  $c$  copies of the group  $C$ ); the set of its generators has inner measure 0 and outer measure 1. To see this, we call a subset of that group a  $c$ -set if it is (in an obvious sense) a cylinder set over a countable number of coördinates. It has been shown to us by S. Kakutani that the inner measure of a subset  $S$  of our group is equal to the supremum of the measures of measurable  $c$ -sets contained in  $S$ . Now it is clear that an element of the group is a generator if and only if its coördinates are linearly independent elements of  $\overline{C}$ . Therefore every  $c$ -set contains elements which are not generators; and this means that the inner measure of the set of generators is 0. On the other hand it is easily seen that every  $c$ -set of positive measure contains a generator (using Theorem III for the special case,  $G =$  direct sum of countably many copies of  $C$ ); hence the complement of the set of generators is of inner measure 0, or the outer measure of the set of generators is 1.

The situation for totally disconnected monothetic groups is this. The set of generators of a group which is cyclic of order  $p^n$ , or  $p$ -adic, has measure  $1 - 1/p$ . From this and the structure of an arbitrary totally disconnected monothetic group  $Z$  (§ 4) one concludes easily that the measure of the set of generators of  $Z$  is equal to  $\prod (1 - 1/p)$ , extended over those primes  $p$  for which  $Z_p \neq 0$ , and hence that this measure may take any value between 0 and 1.

<sup>3</sup> Van Dantzig, D., "Zur topologischen Algebra," *Mathematische Annalen*, 107, 591 (1933).



<sup>1</sup> Weil, André, "L'intégration dans les groupes topologiques et ses applications," Paris, p. 97 (1938).

<sup>2</sup> This theorem was stated by Schreier, J., and Ulam, S., "Sur le nombre des générateurs d'un groupe topologique compact et connexe," *Fund. Math.*, 24, 304 (1935).

## ON THE DIRICHLET PROBLEM FOR THE HYPERBOLIC CASE

By J. HADAMARD

NEW YORK CITY

Communicated March 31, 1942

After having ascertained that Cauchy's problem admits of no solution in general, for equations of the elliptic type, it is natural to investigate whether, conversely, such is the case as concerns Dirichlet's problem for hyperbolic equations: to begin, for any hyperbolic equation in two independent variables

$$\frac{\partial^2 u}{\partial x \partial y} + A(x, y) \frac{\partial u}{\partial x} + B \frac{\partial u}{\partial y} + Cu = 0. \quad (E)$$

The question is to determine a solution  $u$  of such an equation, being given the values of  $u$  along a closed contour  $C$ . It is assumed, for simplicity, that  $C$  is convex with respect to both characteristic directions, i.e., met at no more than two points by any parallel to the  $x$ - or the  $y$ -axis (unless it contains a segment of such a parallel) so that either  $x$  or  $y$  has only one maximum and one minimum on  $C$ . Points where such a maximum or minimum takes place will be called "vertices" of  $C$ , a "side" being the segment of  $C$  between two consecutive vertices.

I have dealt with that question from 1921 on.<sup>1</sup> It can be easily seen that such a Dirichlet problem admits of no solution when  $C$  has only two or three vertices, or, also, when it consists of a rectangle  $MNPQ$  with its sides parallel to the axes. But in the case of four vertices—e.g., when  $C$  is an ellipse—the problem had hitherto been treated only for the simplest equation

$$\frac{\partial^2 u}{\partial x \partial y} = 0, \quad (e)$$

where discussion has been carried out thanks to the fact that, for any solution of (e) and for any rectangle such as  $MNPQ$ , we must have

$$u_M + u_P = u_N + u_Q, \quad (1)$$

this leading to the conclusion that the problem is, in general, again im-

possible, the conditions of possibility depending, in an unforeseen manner, on Diophantine properties of the contour.<sup>2</sup>

Such are the results as concerns (e). As we said, a great part of them is related to the identity (1), the analogue of which does not exist for other equations of the type (E), even for the simplest class of them after (e), i.e., equations which are integrated by an immediate application of Laplace's method. Attempts to generalize them to such equations remained unsuccessful until, recently, I have discovered that the conclusions are of a quite different nature.

1. The general solution of an equation which is integrated by an immediate application of Laplace's method can be, by means of an insignificant change of unknown, written<sup>3</sup>

$$u = X + \int \lambda Y dy, \quad (2)$$

where  $X$  denotes an arbitrary function of  $x$ ;  $Y$ , an arbitrary function of  $y$ ;  $\lambda = \lambda(x, y)$ , a determinate function of  $x$  and  $y$ , the equation itself being

$$\frac{\partial^2 u}{\partial x \partial y} - \frac{1}{\lambda} \frac{\partial \lambda}{\partial x} \frac{\partial u}{\partial y} = 0. \quad (e')$$

$\lambda$  must not vanish inside  $C$  if we want the coefficients of the equation to be regular.

We shall limit ourselves to contours  $C$  admitting the  $y$ -axis as an axis of symmetry, so that the two ordinates  $\bar{y}$  and  $\underline{y} < \bar{y}$  which correspond to the same value of  $x$  are even functions of  $x$ , the maximum of  $\bar{y}$  and the minimum of  $\underline{y}$  occurring for  $x = 0$ .

If  $\lambda$  simply consists of a product  $f(x)g(y)$ , the equation is reducible to (e) by changing  $u$  into  $u/f(x)$ . We begin by taking the next case, in which  $\lambda$  is a sum of two such terms, which case may generally be reconducted to assuming  $\lambda = x + \varphi(y)$ .

Now, we are given, for every value of  $x$  between  $-a$  and  $+a$ , not only two values  $\bar{y}$  and  $\underline{y}$  of  $y$ , but two corresponding values  $\bar{u}$  and  $\underline{u}$  of  $u$  and we must have

$$X + \int_0^{\bar{y}} \lambda Y dy = \bar{u}, \quad (3)$$

$$X + \int_0^{\underline{y}} \lambda Y dy = \underline{u}. \quad (3')$$

These, by subtracting and writing  $(u)$  for  $\bar{u} - \underline{u}$ ,

$$\int_{\underline{y}}^{\bar{y}} \lambda Y dy = (u). \quad (4)$$

Conversely, it will be sufficient to satisfy the latter condition: for, then, we can choose  $X$  by (3) and it will satisfy (3').



Moreover, we shall consider simultaneously opposite values  $x$  and  $-x$  of the abscissa and combine the corresponding equations (4) by addition and subtraction. If we write, for every  $x$  between 0 and  $a$

$$\begin{aligned}(u)(x) + (u)(-x) &= 2v(x), \\ (u)(x) - (u)(-x) &= 2xw(x),\end{aligned}\tag{5}$$

we shall have, for  $\lambda = x + \varphi(y)$ ,

$$\int_{\underline{y}}^{\bar{y}} Y dy = w(x), \quad \int_{\underline{y}}^{\bar{y}} \varphi(y) Y dy = v(x).$$

These equations are both satisfied for  $x = a$ , each member being zero, so that we can replace them by their derivatives with respect to  $x$ , viz.,

$$\bar{y}'\bar{Y} - \underline{y}'\underline{Y} = dw/dx, \quad \varphi(\bar{y})\bar{y}'\bar{Y} - \varphi(\underline{y})\underline{y}'\underline{Y} = dv/dx \tag{6}$$

$\bar{Y}$  and  $\underline{Y}$  standing for  $Y(\bar{y})$ ,  $Y(\underline{y})$ ; or, solving with respect to  $\bar{Y}$ ,  $\underline{Y}$ ,

$$\bar{Y} = \frac{dv/d\bar{y} - \varphi(\underline{y})dw/d\bar{y}}{\varphi - \underline{\varphi}} \tag{6'}$$

with a similar formula for  $\underline{Y}$ . If we assume  $\varphi$  to be a monotone function of  $y$  (or even under more general circumstances), the denominator will not vanish except as the vertices  $x = \pm a$ . If  $\varphi_0$  denotes the value of  $\varphi$  at one of these vertices, we have a condition of possibility, viz.,

$$dv/dy - \varphi_0 dw/dy = 0, \tag{7}$$

which is common to (6') and the formula for  $\underline{Y}$ , at the aforesaid vertices. This will be sufficient in order that  $\bar{Y}$  and  $\underline{Y}$  exist, if (1) at such a vertex,  $\varphi$  has a derivative different from zero; (2) the left-hand member of (7) admits of a derivative. But a second condition of possibility is required if we want  $\bar{Y}$  and  $\underline{Y}$  to be equal (in other words,  $Y$  not be affected with a discontinuity of the first kind).

At the vertices on the  $y$ -axis,  $\bar{Y}$  and  $\underline{Y}$  will, on the contrary, become infinite (excepting, however, the hypothesis of angular points at those vertices) because such is the case for  $dx/d\bar{y}$  or  $dx/d\underline{y}$ ; but  $\int \frac{dx}{d\bar{y}} d\bar{y}$  and  $\int \frac{dx}{d\underline{y}} d\underline{y}$  being absolutely convergent, this will not prevent  $u$  from being continuous around these points.

Such a result is of an unexpected simplicity, implying one single condition or two conditions of possibility, in contrast with the (in general, transcendental) system of conditions which intervenes when we deal with equation (e).

2. For arbitrary forms of  $\lambda$ , the ordinary equations (6) will be replaced

by integral equations of the Volterra type. To simplify scripture, we shall now assume  $C$  to be symmetrical with respect not only to the  $y$ -, but also to the  $x$ -axis, so that  $\underline{y} = -\bar{y}$ : a, strictly speaking, unnecessary assumption, but which does not diminish generality, as is easily seen by means of a transformation carried out on the coördinate  $y$ . Thus (4) can be written

$$\int_0^{\bar{y}} [\lambda(x, y) \bar{Y} + \lambda(x, -y) \underline{Y}] dy = (u),$$

now implying two unknown functions  $\bar{Y}$ ,  $\underline{Y}$  in the interval  $(0, b)$ , where  $b$  is the maximum ordinate on  $C$ . As above, we write it for opposite values of  $x$  and combine by addition and subtraction, getting

$$\int_0^{\bar{y}} [\beta(x, y) \bar{Y} + \beta(x, -y) \underline{Y}] dy = v, \quad \int_0^{\bar{y}} [\gamma(x, y) \bar{Y} + \gamma(x, -y) \underline{Y}] dy = w$$

where

$$\lambda(x, y) + \lambda(-x, y) = \beta(x, y), \quad \lambda(x, y) - \lambda(-x, y) = 2x\gamma(x, y),$$

while  $v, w$  still keep their meanings (5). These are two Volterra equations of the first kind, the integrands depending on  $y$  by the intermediary of  $x$ . As classical, we reconduct them to the second kind by differentiation (the equations being evidently satisfied for  $y = 0$ ) and we get

$$\begin{aligned} \bar{\beta} \bar{Y} + \underline{\beta} \underline{Y} + \frac{dx}{d\bar{y}} \int_0^{\bar{y}} \left[ \frac{\partial \beta}{\partial x}(x, y) \bar{Y} + \frac{\partial \beta}{\partial x}(x, -y) \underline{Y} \right] dy &= \frac{dv}{d\bar{y}}, \\ \bar{\gamma} \bar{Y} + \underline{\gamma} \underline{Y} + \frac{dx}{d\bar{y}} \int_0^{\bar{y}} \left[ \frac{\partial \gamma}{\partial x}(x, y) \bar{Y} + \frac{\partial \gamma}{\partial x}(x, -y) \underline{Y} \right] dy &= \frac{dw}{d\bar{y}} \end{aligned}$$

$\bar{\beta}, \underline{\beta}, \bar{\gamma}, \underline{\gamma}$  standing for  $\beta(x, \bar{y}), \beta(x, -\bar{y}), \gamma(x, \bar{y}), \gamma(x, -\bar{y})$ .

We again assume that the determinant

$$D = \bar{\beta} \underline{\gamma} - \underline{\beta} \bar{\gamma}$$

is different from zero except at the vertices (for instance, that  $\beta/\gamma$  is a monotone function of  $y$  along the contour). At the vertices on the  $x$ -axis,  $D$  vanishes necessarily and we have the condition of possibility

$$\gamma_0 \frac{dv}{d\bar{y}} - \beta_0 \frac{dw}{d\bar{y}} = 0 \quad (7')$$

which is the analogue of (7),  $\beta_0$  and  $\gamma_0$  being the values of  $\beta$  and  $\gamma$  for  $x = a, y = 0$ . This will, conversely, prove sufficient (except for the continuity of  $\bar{Y}$  and  $\underline{Y}$ ), if  $D$  is only of the first order in  $y$  (that is, if  $\beta \frac{d\gamma}{d\bar{y}} - \gamma \frac{d\beta}{d\bar{y}} \neq 0$  at the vertex) and if, moreover, the left-hand member of (7') has a derivative: for  $dx/d\bar{y}$  is, as we can assume, at least of the order of  $y$  in the

vicinity of  $y = 0$  and, therefore, the product  $D^{-1} \frac{dx}{d\bar{y}}$  will remain regular for  $y = 0$ , so that no singularity will occur except, generally, a discontinuity of the first kind. For any positive  $y$  less than  $b$ , such equations will be solved by the classical formulae of Volterra. For  $y = b$ , there will be again no difficulty if the corresponding vertex is an angular point. If there is a tangent parallel to the  $x$ -axis,  $dx/d\bar{y}$  will become infinite<sup>4</sup> and the same takes place, as a rule, for the solutions  $\bar{Y}$ ,  $\underline{Y}$ , as already appears on the special case treated in the preceding section; but, if we introduce the function

$$\xi(\bar{y}) = \frac{1}{\bar{y}} \frac{dx}{d\bar{y}}$$

and the new unknowns  $\bar{Z}$ ,  $\underline{Z}$  defined by

$$\bar{Y}(\bar{y}) = \xi(\bar{y})\bar{Z}(\bar{y}), \quad \underline{Y}(\bar{y}) = \xi(\bar{y})\underline{Z}(\bar{y}),$$

we shall now see that  $\bar{Z}$ ,  $\underline{Z}$  will remain finite, so that, as in the preceding section,  $u$  will remain continuous. The equations for  $\bar{Z}$ ,  $\underline{Z}$  will be of the form

$$\begin{aligned} \bar{Z}(\bar{y}) + \int_0^{\bar{y}} \xi(y) [H\bar{Z}(y) + J\underline{Z}(y)] dy &= \frac{\bar{y}}{D} \left( \gamma \frac{dv}{dx} - \beta \frac{dw}{dx} \right) \\ \underline{Z}(\bar{y}) + \int_0^{\bar{y}} \xi(y) [K\bar{Z}(y) + L\underline{Z}(y)] dy &= \frac{\bar{y}}{D} \left( \beta \frac{dw}{dx} - \gamma \frac{dv}{dx} \right) \end{aligned} \quad (8)$$

( $H, J, K, L$ , finite functions of  $y, \bar{y}$ ). For any shape of  $C$  in the vicinity of the vertex (under our general assumptions)  $\xi(y)$  will be such that the integral  $\int \xi(y) dy$  converges. Therefore, in the Volterra series which solve (8), each of the successive terms

$$\begin{aligned} \bar{Z}_h &= - \int_0^{\bar{y}} \xi(y) [H\bar{Z}_{h-1}(y) + J\underline{Z}_{h-1}(y)] dy \\ \underline{Z}_h &= - \int_0^{\bar{y}} \xi(y) [K\bar{Z}_{h-1}(y) + L\underline{Z}_{h-1}(y)] dy \end{aligned} \quad (9)$$

will remain finite even for  $\bar{y} = b$ . Let us write

$$\bar{Z}_h = \bar{A}_h \bar{y}^h, \quad \underline{Z}_h = \underline{A}_h \bar{y}^h, \quad |\bar{A}_h|, |\underline{A}_h| \leq M_h.$$

If  $\xi(y)$  remains finite, it is elementary that  $M_h/M_{h-1}$  would tend to zero like  $1/h$ . But at any rate, we have

$$M_h \bar{y}^h \leq k M_{h-1} \int_0^1 \xi(y) y^{h-1} dy \quad (k, \text{ a positive constant})$$

Let us assume that, for any  $y, y_1$  such that  $0 < y_1 < y$ , the ratio  $\xi(y_1)/\xi(y)$  remains less than a fixed number  $l$ : this is certainly true if  $C$  is con-

vex in the ordinary meaning of the word (so that  $|dx/dy|$  is an increasing function of  $y$ ) and  $x$  is of the second order for  $y = 0$ . Let us also, by a change of units, take  $b = 1$ . The integral

$$I_h = \int_0^1 \xi(y) y^{h-1} dy$$

tends to zero when  $h$  increases to infinity. Then, the analogous integral with the upper limit  $\bar{y}$  can be written

$$\int_0^{\bar{y}} \xi(y) y^{h-1} dy = \bar{y}^h \int_0^1 \xi(t\bar{y}) t^{h-1} dt \leq \bar{y}^h I_h \cdot l,$$

on account of the assumption concerning  $\xi(y)$ . Therefore, we can take

$$M_h \leq kl M_{h-1} I_h,$$

where the last factor tends to zero,<sup>5</sup> and we see that the series for  $\bar{Z}$ ,  $\underline{Z}$  are uniformly convergent for every  $y$  between 0 and 1, limits included.

3. The results are, as we see, mainly the same as in the preceding section. But they are in full discrepancy with what has been found for (e)—and also with what happens for any equation of the type (É) whenever the contour has only two or three vertices—as only one or two conditions of possibility are required, at the vertices  $y = 0$ ,  $x = \pm a$ . Which of these so different conclusions gives us a right idea of what occurs in the general case? Perhaps neither of them: Poincaré has taught us to diffide of conclusions drawn, as to a general problem, from the special cases the treatment of which happens to be within our reach.

<sup>1</sup> *Proc. Benares Math. Soc.*, 3, 39 (1921); *Lectures on Cauchy's Problem*, French edition, Paris, 1932, Appendix II; "Conferences Internationales des Sciences Math., Geneve, 1935," published in *L'Enseignement Mathématique*, Paris-Geneve, 1936; *Jour. Chinese Math. Soc.*, 2, 6-20 (1937).

<sup>2</sup> Besides the papers mentioned in reference 1, above, see Huber, A., *Monatshefte Math. Phys.*, 39 (1932); Bourgin, D. G., and Duffin, R., *Bull. Am. Math. Soc.*, 45 (1939); and a beautiful treatment by John, F., *Am. J. Math.*, 63 (1941).

<sup>3</sup> The general solution of an equation admitting immediate integration by Laplace's method is written in Darboux's *Leçons*, vol. 2,

$$\alpha(X + \int \beta Y dy).$$

Both  $\alpha$  and  $\beta$  being necessarily different from zero inside  $C$ , we do not diminish generality by taking  $\alpha = 1$ , and we have written  $\lambda$  instead of  $\beta$ .

<sup>4</sup> Such will be also the case for the right-hand members of the equations which will similarly become infinite like  $\xi$ . Near  $(0, b)$ , it is natural to assume  $v, w$  to be regular in  $x$ , but not, in general, in  $y$ .

<sup>5</sup> If, in the vicinity of the vertex on the  $y$ -axis,  $1 - \bar{y}$  is of the order of  $x^m$ , with  $m$  being any constant greater than one,  $I_h$  will be  $1/m B(h, 1/m)$ .  $B$ , an eulerian integral of the first kind, is asymptotically equal to  $\Gamma(1/m) h^{-1/m}$ , so that the series (9) will converge like  $\Sigma (h'y)^h: (h!)^{1/m}$ .



# PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

Volume 28

July 15, 1942

Number

*Copyright 1942 by the National Academy of Sciences*

---

## CHROMOSOMES OF THE RED FOX\*

BY LOUISE WIPF AND RICHARD M. SHACKELFORD

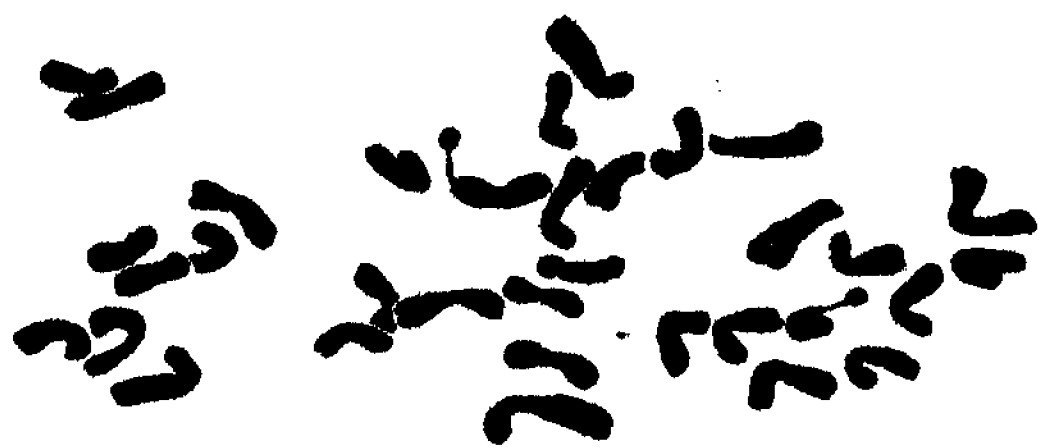
DEPARTMENTS OF VETERINARY SCIENCE AND GENETICS, UNIVERSITY OF WISCONSIN

Communicated June 11, 1942

A number of mutant color types have arisen, directly or indirectly, from the red fox (*Vulpes vulpes* L.). Two different mutations to black [standard black (silver) and Alaskan black (silver)] had occurred in wild red foxes of North America prior to ranch breeding. The platinum character is the result of a mutation from ranch-bred standard black (silver) foxes. The present study is concerned with both the number and the comparative forms of the chromosomes in red, standard black (silver) and platinum-silver foxes. Wodsdalek<sup>1</sup> reported 42 as the chromosome number of the male red fox. Later Andres<sup>2</sup> found the number to be 34.

Testes were collected during the breeding season of 1942 in order to obtain meiotic and mitotic divisions. A red and a black (silver) fox were killed by electrocution, and another red fox by an injection of ether into the heart. The material from the platinum-silver fox was obtained by castration. The testes in each case were immediately excised and lacerated, and the seminiferous tubules were teased out and placed in Carnoy's alcohol-acetic acid-chloroform solution (7:2:1). The complete operation required less than five minutes. The tubules were allowed to remain in the fixative from 30 minutes to one hour; they were then transferred to 95 per cent alcohol for an hour and stored in 80 per cent alcohol. Temporary mounts prepared by the aceto-carmines smear method provided excellent material for study.

Polar views of diploid equatorial plates from seminiferous tubules show 34 chromosomes in the red, black (silver) and platinum-silver foxes (Figs. 1, 2, 3). The chromosomes of all three color types appear similar in size, shape and other morphological characteristics. Longitudinal splits and median or sub-terminal spindle fibre attachment regions are common (Fig. 3). The homologues of several chromosome pairs are easily identified. The chromosomes vary in length, the shortest being about one-half that of the longest. A pair of satellite chromosomes is present in each case.



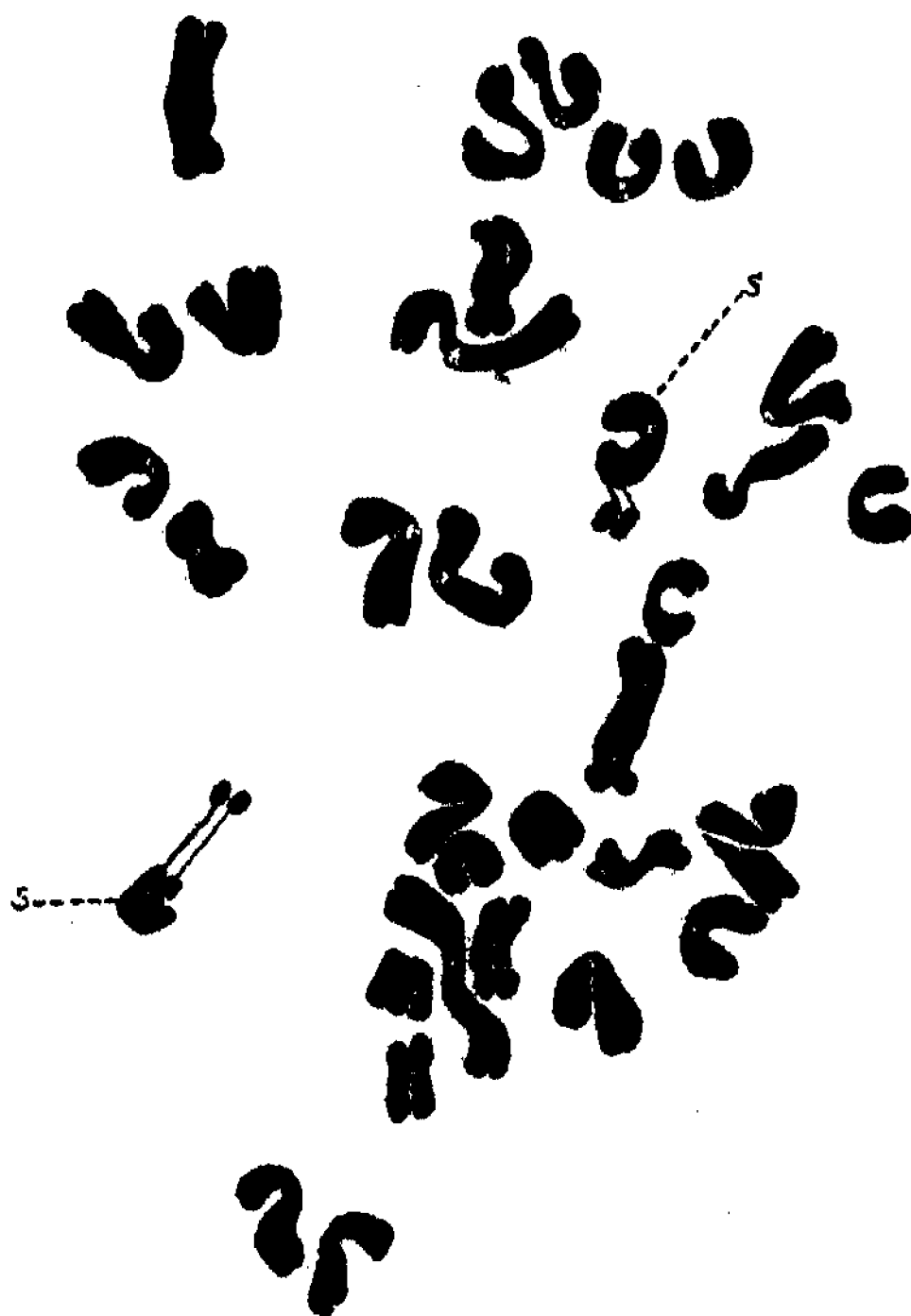
1



2



4



3



5

EXPLANATION ON OPPOSITE PAGE

These satellite chromosomes are of medium size and have median spindle-fibre attachment regions (Fig. 3 (s)). As a general rule, satellite chromosomes of both plants and animals are found associated with the nucleolus. This may be true in the fox, for a resting nucleus shows chromatic material attached to the nucleolus in two places (Fig. 4).

On the first meiotic equatorial plate the chromosomes are shorter, thicker and more closely packed than in the mitotic divisions; hence the determination of the haploid number is somewhat difficult. However, 17 pairs (Fig. 5), including one pair with attached satellites, are present on the heterotypic equatorial plate. Spindle-fibre attachment regions are easily recognized at this stage. There appears to be more variation in the size of the chromosomes at meiosis than in mitosis.

The occurrence of satellite chromosomes in the fox is of special interest, since, so far as the writers are aware, they have not been previously recorded in mammals. Coonen,<sup>3</sup> in a review of the literature on satellite chromosomes in plants and animals, found them common in plants but comparatively rare in animals. He cites reports of only six animals in which satellites have been noted: the mosquito, *Drosophila*, *Bibio*, *Amblystoma*, *Opalina* and *Salmo*. The present observations agree with those of Andres<sup>2</sup> as to the chromosome number in the red fox, but he makes no reference to satellite chromosomes. His figure 1 shows 34 chromosomes plus a small chromosome for which he gives no explanation. Since Andres<sup>2</sup> used sectioned material instead of smears it may be that the satellite connection was lost or obscured by the method used, and that the extra chromatic particle actually was the satellite.

The platinum-silver character in foxes has been reported<sup>4</sup> as lethal in the homozygous condition. It would be of interest to investigate the possibilities of observable differences in the chromosomes of foxes heterozygous for this character. It is not to be expected that minor differences would be seen at the equatorial plate, but further search may possibly reveal dissimilarities in the pachytene stage.

The writers are indebted to Professor W. Wisnicky, Director of Fur Farm Research, and Professor L. J. Cole for materials and suggestions.

#### EXPLANATION OF PLATE

Figures 1-3. Equatorial plates from seminiferous tubules of the red fox, figure 1,  $\times 2140$ ; black (silver), figure 2,  $\times 2140$ ; platinum-silver, figure 3,  $\times 2920$ ; showing the diploid chromosome number. Thirty-four chromosomes are present in each case. Figure 3 (s), satellite chromosome.

Figure 4. Resting nucleus from seminiferous tubule, red fox, showing chromatic attachments to the nucleole.  $\times 2920$ .

Figure 5. Equatorial plate, first meiotic division, from seminiferous tubules of black (silver) fox.  $\times 2920$ .



Further acknowledgment is extended to Professors D. C. Cooper and C. E. Allen for advice and constructive criticisms.

\* Joint contribution from the Departments of Veterinary Science and Genetics (Paper No. 284), Agricultural Experiment Station. Published with the approval of the Director of the Station.

<sup>1</sup> Wodsdalek, J. E., *Anat. Rec.*, **51**, 70 (1931).

<sup>2</sup> Andres, A. H., *Cytologia*, **9**, 35-37 (1938).

<sup>3</sup> Coonen, L. P., *Amer. Jour. Bot.*, **26**, 49-58 (1939).

<sup>4</sup> Cole, L. J., and Shackelford, R. M. (in press).

## SALIVARY GLAND TYPE CHROMOSOMES IN MOSQUITOES

BY EILEEN SUTTON

CARNEGIE INSTITUTION OF WASHINGTON, COLD SPRING HARBOR, N. Y.

Communicated June 15, 1942

Mosquitoes have been the subject of several cytological studies, but in these insects very little consideration has been given to the type of chromosome characteristic of the salivary glands and other organs of many Dipteran larvae. In recent work, chromosomes from the large larval mid-gut cells of mosquitoes have been described<sup>1,2</sup> as similar to the salivary gland type, but lacking the regular alternation of bands and achromatic regions and consisting merely of a linear series of chromatic masses.

Giant chromosomes with regular banded structure may be obtained, however, from certain tissues in the later stages of development. These chromosomes will be described here. The most satisfactory preparations were obtained from the Malpighian tubes of the imago, pupa or fourth instar larva, but such chromosomes were also found in the salivary glands, gastric caeca and mid-gut of the prepupal stage.

### DESCRIPTION OF PLATE

Figure 1. Nucleus from Malpighian tubes of *Culex pipiens* male, pupal stage. Ca.  $\times 510$ .

Figure 2. Chromosome from *C. pipiens* male, with nucleolus. Ca.  $\times 1160$ .

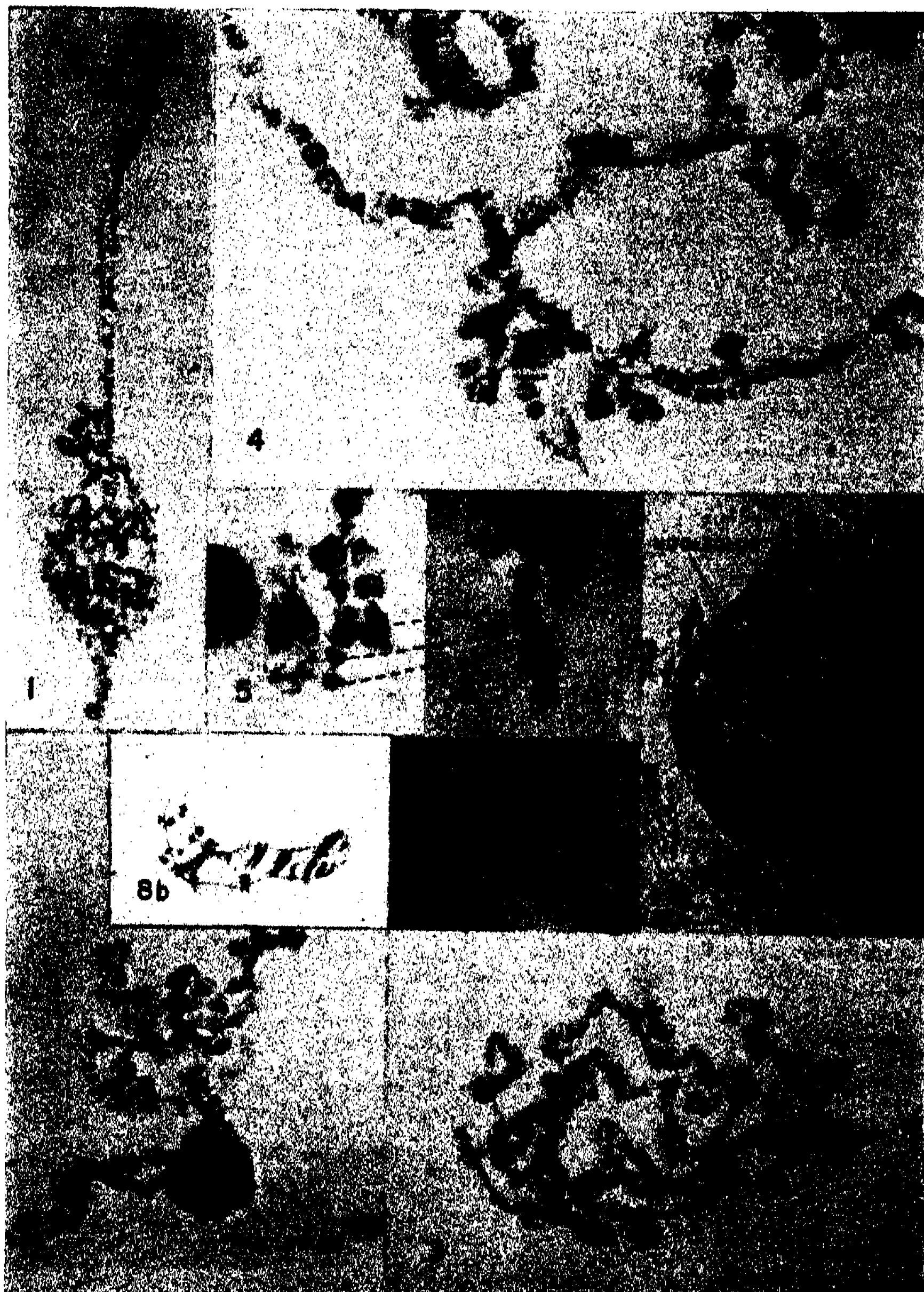
Figure 3. Nucleus from *Aedes aegypti* female. Ca.  $\times 580$ .

Figure 4. Part of nucleus from *A. aegypti* male showing banded regions and "weak spots." Ca.  $\times 1160$ .

Figures 5 and 6. Comparison of the same region in two different cells of *A. aegypti*, fourth instar larva. Ca.  $\times 1160$ .

Figure 7. Nucleus from *A. aegypti*, fourth instar larva, showing globular structure and nucleolus. Ca.  $\times 580$ .

Figure 8 (a). Section of a chromosome from *A. aegypti* pupa, with peculiar banded structure. Ca.  $\times 1160$ . (b) Camera lucida drawing of same.



DESCRIPTION ON OPPOSITE PAGE

In the salivary glands of the adult mosquito the chromosomes are apparently of the normal somatic type.

*Technique.*—Eggs of the species *Culex pipiens* and *Aedes aegypti*<sup>3</sup> were raised in tap-water at room temperature and the larvae were fed with breadcrumbs. The best slides were obtained from individuals (fourth instar larvae, pupae or newly emerged adults) which had been kept at low temperature (10–18°C.) for two or three days before killing.

The standard technique for smearing salivary gland chromosomes (by staining for a few minutes in aceto-carmin or acetic orcein placing on the slide in a drop of stain and flattening by pressure on the coverslip) was found to be ineffective for the mosquito chromosomes. These chromosomes are very soft, and instead of spreading them out, the standard procedure merely squashes the whole nucleus.

The best treatment that I have found so far is the following modification of the standard technique: The tissue is dissected out and placed for one minute in acetic alcohol. It is then stained in acetic orcein (1% orcein in 45% acetic acid) for about one hour, and smeared in the usual way.

*The Salivary Type Chromosomes.*—(1) *Culex pipiens*. Figure 1 in the plate shows a whole nucleus from a smear of the Malpighian tubes of *Culex pipiens*. One chromosome pair lies free, and the banded structure can be clearly seen. In some preparations I have seen the three pairs of chromosomes ( $2n = 6$ ), synapsed along most of their length, and completely separated from each other, with no apparent chromocenter. Regions which appear heterochromatic like the chromocenter regions in *Drosophila melanogaster* have not been observed. A nucleolus is associated with one pair of chromosomes near the distal end of one of the arms (Fig. 2).

No detailed study of the banding pattern has been attempted, but it is clear that a map could be made for each chromosome, similar to the salivary gland chromosome maps of *Drosophila*.

In a cursory study, no difference was detected between the chromosome complexes of males and females. It may be noted that the sexes are not certainly distinguishable by their chromosomes in normal mitotic divisions.<sup>4</sup>

(2) *Aedes aegypti*. Figures 3 and 4 show a whole nucleus of *Aedes aegypti* and part of a nucleus at a higher magnification. The banded structure is clear in some places, but the pattern of banding cannot be followed along individual chromosome arms, because it is interrupted by "weak spots." At these places the chromosomes often fragment in smearing, and parts of the same or of different chromosomes adhere in these regions, so that the continuity of any particular arm is obscured. These contact points can be seen in figure 4. Parts of chromosomes are very often connected by attenuated threads.

As in *Culex pipiens* there is no single chromocenter such as is found in

salivary gland nuclei of *Drosophila*. It is possible that the contacts are due to non-specific pairing between heterochromatic segments which are interspersed between the banded regions all along the chromosome arms.

Figures 5 and 6 show comparable banded fragments from different nuclei. By a series of such comparisons it would be possible to map the banding pattern for short regions, but it is doubtful whether the sequence for a whole chromosome could be worked out.

Although a nucleolus is present earlier in the development of the nuclei (Fig. 7), it is not visible at later stages after treatment with acetic orcein, in marked contrast to the nucleolus of *Culex pipiens*.

*Development of the Salivary Gland Type Chromosomes.*—The development of the fully banded chromosomes in mosquitoes has not been studied to any considerable extent, but a few observations, made in the course of dissecting tissues from individuals at various stages of development, may be of some interest. In this respect the two species studied are similar.

The regular banding is not present in chromosomes at the earlier stages (up to third instar larvae), nor do all chromosomes in the Malpighian tubes of later stages necessarily attain this type. In young larvae all, and in later stages some of the cells in question have chromosomes in which the nucleic acid is accumulated in more or less spherical masses of different sizes, separated by achromatic regions (Fig. 7).

Sometimes the globules appear to be connected in a linear series only by fine chromatic threads. In a few figures, short sections of chromosome showed a peculiar banded structure, in which the bands appeared to be connected by paired threads (Fig. 8).

It seems that the dense accumulations of nucleic acid in the globule stage are redistributed (perhaps spread out laterally, or perhaps partially lost) in the formation of the banded type of chromosome. Some intermediate cells are found in the later stages, where well-defined bands are seen in some sections of the chromosomes, while other sections approximate to the globule type.

Tentatively, a comparison may be drawn between this shift in nucleic acid attachment and the experimental results obtained by Kodani.<sup>5</sup> The globule stage described here bears some resemblance to the  $\gamma$  1 stage in his treated chromosomes. It must be remembered, however, that while the transition in mosquito is a normal developmental process, the apparent reversal of this process obtained by Kodani is induced by somewhat drastic treatment, and it is not yet known whether this induced effect is reversible.

*Summary.*—A brief description is presented of the development and characteristics of salivary gland type chromosomes in two mosquito species (*Culex pipiens* and *Aedes aegypti*) in both of which the diploid chromosome number is six. The differences observed between the two species indicate

that these chromosomes might be useful as a diagnostic character in distinguishing different species of mosquito.

<sup>1</sup> Berger, C. A., *Carnegie Inst. Pub.* 496, *Contrib. Embryo.*, **167**, 211-232 (1938).

<sup>2</sup> Bogojawlensky, K. S., *Zeit. f. Zellforsch u. Mikr. Anat.*, **22**, 47-53 (1934).

<sup>3</sup> Obtained by courtesy of Dr. J. Maier, Rockefeller Inst., New York.

<sup>4</sup> Whiting, P. W., *Jour. Morph.*, **28**, 523-577 (1917).

<sup>5</sup> Kodani, M., *Jour. Hered.*, **33**, 115-133 (1942).

---

## PURINES AS GROWTH REQUIREMENTS OF *SPIRILLUM SERPENS*

BY DERROL E. PENNINGTON\*

DEPARTMENT OF CHEMISTRY, UNIVERSITY OF TEXAS

Communicated May 27, 1942

The recognition of the purine and pyrimidine bases as factors affecting the growth of microorganisms has been reported with increasing frequency during the past few years. Richardson<sup>1</sup> in 1936 found that under certain conditions uracil was essential for the growth of *Staphylococcus aureus*. Möller<sup>2</sup> showed that adenine was required for the growth of *Streptobacterium plantarum*, while Pappenheimer and Hottle<sup>3</sup> found adenine to be necessary for the growth of a strain of group A hemolytic streptococci. In the latter case adenine could be replaced by hypoxanthine, guanine, xanthine, guanylic acid or adenylic acid. Furthermore, it was observed that purines were not required by this organism if the carbon dioxide tension above the medium was maintained sufficiently high. Snell and Mitchell<sup>4</sup> have reported the requirements of several lactic acid bacteria for purine and pyrimidine bases. *Streptococcus lactis* was found to require adenine and thymine for growth. Guanine was found to be essential for the growth of *Leuconostoc mesenteroides*, while uracil stimulated the growth of the latter organism and of *Lactobacillus arabinosus*. Robbins and Kavanagh<sup>5,6</sup> have recently described the effect of guanine and hypoxanthine on the growth of *Phycomyces*. In the present communication we wish to report the essential nature of several purine bases for the growth of a strain of *Spirillum serpens*.

Investigation of the cultural requirements of this organism (carried by American Type Culture Collection as No. 8084) showed that although it could not be grown in simple synthetic media the addition to such media of small amounts of natural extracts caused growth to occur. The active principle of such extracts was isolated and was found to consist of purine bases. Hypoxanthine, adenine and guanine were found to affect the growth of the organism.

*Experimental.*—The basal medium used had the following composition:

	AMOUNT PER LITER OF MEDIUM
Asparagine.....	5.0 g.
Inorganic salts:	
$\text{KH}_2\text{PO}_4$ .....	500 mg.
$\text{K}_2\text{HPO}_4$ .....	500 mg.
$\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ .....	200 mg.
$\text{NaCl}$ .....	10 mg.
$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ .....	10 mg.
$\text{MnSO}_4 \cdot 4\text{H}_2\text{O}$ .....	10 mg.

The pH of the medium was adjusted to 7.2. The addition of the proper purine bases to this medium completely met all requirements of the organism. None of the "B-vitamins" was required.

Cultures were grown in 50 ml. Erlenmeyer flasks. Solutions of materials to be tested were pipetted into the flasks and the volume in each was adjusted to 5.0 ml. with distilled water. To each flask was then added 22.0 ml. of the basal medium described above (27 ml. cultures were required to fill the absorption cell used in determining extent of growth). The flasks were plugged with cotton, sterilized by steaming for 15 minutes, cooled and inoculated. Inoculum was prepared from 24 hour cultures of the organism in beef extract-peptone broth. The cells of such a culture were centrifuged out aseptically, resuspended in sterile water, and one drop of the resulting suspension was used to inoculate each flask.

The growth of the organism was measured by quantitatively comparing turbidities in a thermoelectric turbidimeter.<sup>7</sup> Growth was expressed directly as galvanometer readings.

TABLE 1  
HYPOXANTHINE AND GUANINE

(Growth Expressed as Turbidimeter Galvanometer Readings)

$\mu\text{G. HYPOXANTHINE PER}$ 27 ML. CULTURE	$\mu\text{G. GUANINE ADDED PER 27 ML. CULTURE}$		
	0	10	200
0	29.7*	....	..
3	38.2	30.5	28.2
6	48.5	37.8	29.0
12	59.9	60.0	30.0
30	64.0	64.0	33.0
100	65.8	..	..
250	65.8	..	65.0

\* Absorption cell filled with distilled water gave a reading of approximately 25.

The effect of the addition of hypoxanthine to the basal medium is shown in table 1. This compound alone is able to replace all of the growth-promoting activity of natural extracts. In the third and fourth columns of table 1 is shown the effect of the addition of guanine to cultures containing



hypoxanthine. Guanine inhibits the growth stimulated by hypoxanthine, provided the concentration of guanine is greater than that of hypoxanthine. If the amount of hypoxanthine present is in excess of the amount of guanine, the latter factor has no effect on growth.

TABLE 2

## HYPOXANTHINE AND ADENINE

(Growth Expressed as Turbidimeter Galvanometer Readings)

μG. HYPOXANTHINE PER 27 ML. CULTURE	μG. ADENINE ADDED PER 27 ML. CULTURE						
	0	2	5	10	50	100	200
0	28.0	..	..	..	..	..	..
2	33.5	38.3	35.1	29.8	..	..	..
5	41.0	45.8	47.8	39.1	28.0	28.0	..
10	50.5	52.2	54.8	52.5	29.5	28.0	..
16	55.8	..	..	..	..	..	..
100	67.0	..	..	67.0	66.0	51.0	33.8

The effect of the addition of adenine to cultures containing hypoxanthine is shown in table 2. It can be seen that if the concentration of adenine is equal to or less than the concentration of hypoxanthine, growth greater than that produced by the hypoxanthine alone occurs. If the adenine concentration is greater than that of hypoxanthine, toxic effects are observed, and if the ratio of adenine to hypoxanthine becomes large, the physiological activity of the latter compound is entirely masked. It should be noted that when the concentrations of both factors are large, the toxic effect of adenine is apparent at lower adenine-hypoxanthine ratios than at lower concentrations.

Uric acid, xanthine and uracil have no effect on the physiological activity of hypoxanthine.

TABLE 3

## PHYSIOLOGICAL ACTIVITY OF ADENINE AND GUANINE

(Growth Expressed in Turbidimeter Galvanometer Readings)

μG. GUANINE PER 27 ML. CULTURE	μG. ADENINE PER 27 ML. CULTURE						
	0	5	10	20	50	100	250
0	27.0	32.5	31.1	30.0	29.5	29.0	28.7
5	25.9	55.0	56.1	55.0	34.8	30.8	29.4
10	25.8	46.5	58.0	60.5	55.0	34.8	30.0
20	26.0	32.9	56.0	63.0	60.0	53.0	31.0
50	25.8	30.0	36.8	56.7	63.5	60.0	44.1
100	26.0	30.1	31.8	45.0	57.5	54.2	44.8
250	26.0	28.9	30.1	37.9	55.0	53.5	51.0

Table 3 shows the effect of the addition of various combinations of adenine and guanine to the basal medium. Neither adenine nor guanine alone is

active, but a mixture of approximately equal parts of the two is able to supply the requirements of the organism. If either component of the mixture is appreciably in excess of the other, toxic effects are observed, and if the imbalance is extreme, the physiological activity is completely masked. Neither the purine, xanthine nor the pyrimidines, uracil, cytosine or thymine, have any effect on the action of adenine or guanine. Adenosine and yeast adenylic acid cannot take the place of adenine.

*Discussion.*—The growth requirements of *Spirillum serpens* are completely met in a medium composed of asparagine and inorganic salts supplemented with hypoxanthine or an equimolecular mixture of adenine and guanine. The organism thus differs from other organisms which have been reported to require purines in that all of the latter organisms have rather complex requirements in addition to the purines.

The reversible inhibitory action of certain purines on the physiological activity of others has been described. A similar relationship has not hitherto been reported in the case of the purine requirements of microorganisms, but several examples of a similar phenomenon with other nutrilites have been reported. McIlwain<sup>8</sup> showed that pyridine-3-sulfonic acid or its amide interfered with the growth of organisms requiring nicotinic acid. *p*-Aminobenzoic acid in small amounts has been reported to overcome the bactericidal effects of sulfanilamide and its derivatives.<sup>9</sup> Snell<sup>10</sup> has reported that the sulfonic acid analogue of pantothenic acid inhibited the growth of organisms which require this vitamin. The addition of excess pantothenic acid reversed the inhibition.

The toxic effects of the pyridine-3-sulfonic acid, the sulfanilamide compounds and the pantoyl taurine are believed due to the fact that they are structurally very similar to the corresponding naturally occurring nutrilites. The inhibitory substances thus appear to be able to fit into the biochemical patterns normally occupied by the essential nutrilites. They are not, however, able to carry out whatever vital functions are performed by these compounds and the metabolic processes of the organism are blocked at this point.

This picture can be applied in the present instance, but is complicated by the fact that adenine and guanine, which are inhibitory under certain conditions, contribute to the growth of the organism under other conditions. In seeking to rationalize these results the following speculation is suggested.

Assume that the purines function in a physiologically essential complex that contains two purine bases per molecule. (Several important coenzymes are known to have a dinucleotide structure.) Then, according to the results obtained, this complex can function only if (1) both purines are hypoxanthine, (2) one purine is hypoxanthine and the other adenine or (3) one purine is adenine and the other guanine. The complex cannot function physiologically if (1) both purines are adenine, (2) both purines are guanine



or (3) one purine is hypoxanthine and the other guanine. Xanthine and uric acid do not compete for places in the complex. Such an hypothesis is in agreement with all of the observations, and is useful in indicating possible answers to a number of questions. For example, on this assumption it can be seen how it is possible for an excess of adenine or guanine to be toxic while an equimolecular mixture of the two promotes growth, and how it is possible for an excess of adenine to inhibit the action of hypoxanthine while a lower concentration adds to growth.

In line with the above hypothesis it is interesting to recall that Stockstad<sup>11</sup> found *Lactobacillus casei* to require what appeared to be a dinucleotide isolated from liver. This dinucleotide could be partially replaced by a mixture of purine and pyrimidine bases, but only by amounts much greater than required of the dinucleotide.

*Summary.*—Hypoxanthine or an equimolecular mixture of adenine and guanine is essential for the growth of *Spirillum serpens*. Adenine or guanine alone does not support growth, and under certain conditions these compounds are toxic to the organism. No other vitamin-like compounds affect the growth of the organism.

An hypothesis is suggested concerning the toxic effects of adenine and guanine.

The author wishes to thank Dr. R. J. Williams for his advice and encouragement during the course of this work.

\* Standard Brands Fellow for 1941-1942.

<sup>1</sup> Richardson, G. M., *Biochem. Jour.*, **30**, 2185 (1936).

<sup>2</sup> Möller, E. F., *Zeit. physiol. Chem.*, **260**, 246 (1939).

<sup>3</sup> Pappenheimer, A. M., Jr., and Hottel, G. A., *Proc. Soc. Exptl. Biol. Med.*, **44**, 645 (1940).

<sup>4</sup> Snell, E. E., and Mitchell, H. K., *Proc. Nat. Acad. Sci.*, **27**, 1 (1941).

<sup>5</sup> Robbins, W. J., and Kavanagh, F., *Proc. Nat. Acad. Sci.*, **28**, 4 (1942).

<sup>6</sup> Robbins, W. J., and Kavanagh, F., *Ibid.* **28**, 65 (1942).

<sup>7</sup> Williams, R. J., McAlister, E. D., and Roehm, R. R., *Jour. Biol. Chem.*, **83**, 315 (1929).

<sup>8</sup> McIlwain, H., *Nature*, **146**, 653 (1940).

<sup>9</sup> Landy, M., and Wyeno, J., *Proc. Soc. Exptl. Biol. Med.*, **46**, 59 (1941).

<sup>10</sup> Snell, E. E., *Jour. Biol. Chem.*, **141**, 121 (1941).

<sup>11</sup> Stockstad, E. L. R., *Jour. Biol. Chem.*, **139**, 475 (1941).

## THE DISTRIBUTION OF METALLIC ATOMS IN TWO-COMPONENT GLASSES

By JOSEPH S. LUKESH

MINERALOGICAL LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated May 21, 1942

1. *Introduction.*—Most of the present knowledge of glass structure is based on the work of Warren and his school in the interpretation of x-ray diffraction patterns. In the method used by Warren, a diffraction pattern is converted into an intensity curve from which, by application of the Fourier integral, a radial distribution curve is computed. This curve gives an indication of the amount of scattering matter at any distance from any atom. Thus, in the radial distribution curve of silicon dioxide glass, at the shortest interatomic distance, silicon to oxygen =  $1.62 \text{ \AA}$ , there is a well-defined peak. The area under this peak represents four oxygen atoms, an indication that the tetrahedral arrangement of oxygen atoms about silicon which is found in all crystalline silicates obtains in glass. At the next shortest interatomic distance, oxygen to oxygen =  $2.65 \text{ \AA}$ , there is another peak, poorly resolved, but which may be considered to represent approximately six oxygen atoms about each oxygen atom. A third peak occurs at twice the silicon to oxygen separation and may be interpreted as representing the silicon to silicon distance. At larger radial distances, there is so much overlapping of peaks due to the great number of interatomic distances and to the fact that the peaks become broadened because of the random nature of glass structure that further interpretation is difficult.

Similar curves have been computed by Warren and his co-workers for series of glasses consisting of soda-silica,<sup>1</sup> soda-boric oxide,<sup>2</sup> potassium oxide-boric oxide<sup>3</sup> and pure silica and pure boric oxide.<sup>4</sup> In all of these cases only the shortest two or three interatomic distances can be interpreted. No information can be obtained regarding the distribution of the metallic atoms.

2. *The Differential Method.*—Figure 1 (a) shows a typical radial distribution curve for soda-silica glass as obtained by Warren and Biscoe.<sup>1</sup> The peak due to the silicon to oxygen separation,  $1.62 \text{ \AA}$ , is clearly resolved. A second peak occurs at a radial distance of about  $2.4 \text{ \AA}$ . This is poorly resolved and is, actually, two peaks, one due to sodium to oxygen at  $2.35 \text{ \AA}$  and one due to oxygen to oxygen at  $2.65 \text{ \AA}$ . Figure 1 (b) is the curve for pure silica glass, also as obtained by Warren and Biscoe. It occurred to the writer that subtraction of the ordinates of the curve of figure 1 (b) from those of the curve of figure 1 (a) would eliminate all peaks common to both curves and leave only those due to the radial distribution of scattering

matter about sodium, thus providing information about the distribution of the metallic atoms. If this is done, a differential curve, figure 1 (c),

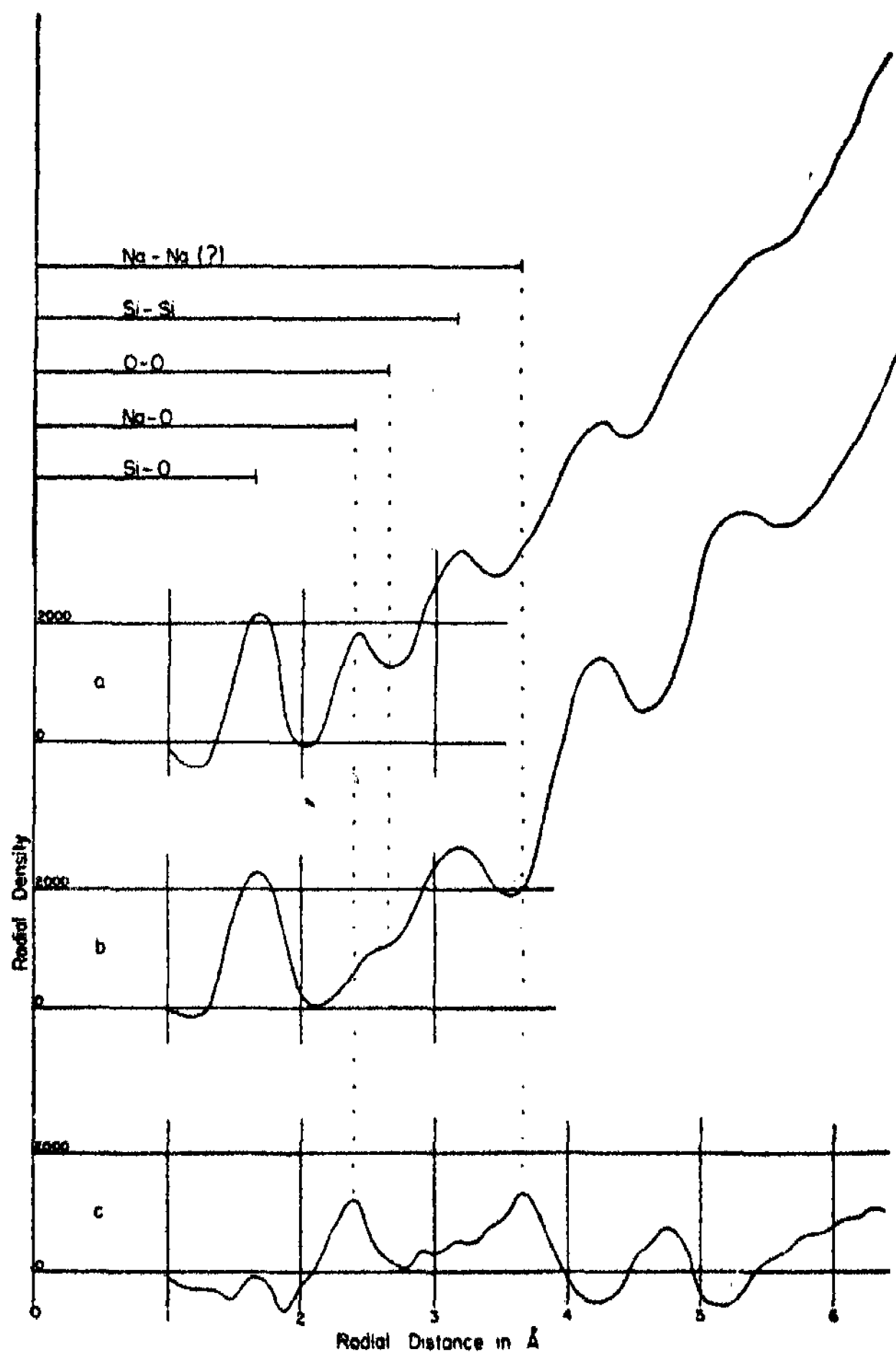


FIGURE 1

(a) Radial distribution curve of 19.5%  $\text{Na}_2\text{O}$ -80.5%  $\text{SiO}_2$  glass (after Warren and Bischoe).

(b) Radial distribution curve of  $\text{SiO}_2$  glass (after Warren and Bischoe).

(c) Differential radial distribution curve obtained by subtraction of silica glass curve from soda-silica glass curve.

results which shows a series of sharp peaks, most of which were not evident in the radial distribution curve of soda-silica glass. There are sharp,

major peaks at distances from the origin of 2.4, 3.65 and 4.75 Å. A fourth major peak occurs in the region of 6.3 Å, but this is less well resolved. These major peaks persist throughout the entire composition range studied by Warren and Biscoe.<sup>1</sup> There are also a number of minor peaks which may or may not be significant, depending upon the extent to which they are affected by errors inherent in the method.

The first major peak on the differential radial distribution curve is due to the sodium-oxygen interatomic distance of 2.35 Å. This peak was observed by Warren and Biscoe<sup>1</sup> on their curve, figure 1 (a), where it appears unresolved from the oxygen-oxygen peak at 2.65 Å. The position of the peak remains substantially unchanged with increasing soda content of the glass, but the area beneath it varies, increasing in general. The second major peak may be attributed to the first sodium-sodium distance. Its position shifts slightly with increasing soda content, and the area beneath it increases regularly, both of which facts are consistent with an interpretation of the peak as being due to a sodium-sodium interatomic distance. The third peak may be due to a second sodium-oxygen distance or a second sodium-sodium distance.

Since the primary purpose of this paper is to outline the principle of the differential method, the above interpretation of the peaks should be considered to be illustrative rather than conclusive. The available data were not designed for this type of treatment, and the results are naturally somewhat crude. It is hoped to supplement this superficial analysis when better data are available.

3. *Sources of Error.*—There are certain sources of error that affect the nature of the differential curve and complicate its interpretation. The data used to construct the curve in figure 1 (c) were taken from published curves which in themselves are accurate only to about one part in ten, and the process of subtraction makes the result subject to considerable error. For this reason, only major peaks which persist throughout the entire composition range can be considered as being significant. Minor peaks may be real, but they may also be due to errors in plotting or measurement.

A theoretical error is introduced by the assumption that the radial distribution of scattering matter about silicon and oxygen does not change. For the radial distances involved here, it seems safe to assume that the interatomic distances do not change appreciably, but the area beneath the curve becomes smaller. In the first place, the area beneath the curve representing the distribution of scattering matter about silicon and oxygen depends upon the total number of silicon and oxygen atoms present in a given volume, and this number decreases as the soda content increases. Also, the number of oxygen atoms about each oxygen becomes less with increasing soda content.<sup>1</sup> The net effect is the subtraction of too much

density. The areas of the residual peaks are too small, and their positions may be shifted slightly.

4. *Sodium and Potassium Borate Glasses.*—Curves similar to figure 1 (c) have been constructed for soda-boric oxide glasses, using the curves of Biscoe and Warren<sup>2</sup> and for potassium oxide-boric oxide glasses, using the curves of Green.<sup>3</sup> The interpretation of these curves is made the more difficult by the fact that the coördination of boron changes from three to four with increasing metal content.<sup>2, 3</sup> The residual peaks are subject to more error than those of silica glasses where the coördination groups of silicon and oxygen do not change. However, in the case of the soda-boric oxide glasses, the differential radial distribution curves show sharp peaks at 3.4 and 4.5 Å which may be due to sodium-sodium distances. In the potassium oxide-boric oxide series, the differential curves show a sharp peak at 3.3 Å which may be the potassium-potassium separation. In both series of glasses, the area under the peaks which are tentatively assigned to the metal to metal separation increases regularly with increasing metal content.

5. *Extension and Improvement of the Differential Method.*—In an accompanying article,<sup>4</sup> M. J. Buerger discusses the application of the differential technique to the synthesis of Fourier series in crystal structure determinations. Applying the idea of atom for atom substitution developed by him to the glass problem, it is possible to eliminate one of the sources of error, namely, the change in area due to the change in coördination of oxygen atoms about oxygen atoms and to the decrease in absolute numbers of silicon and oxygen atoms present in a given volume. The curve of a lithia-silica glass, for instance, might be subtracted from one of a soda-silica glass of the same molal composition. Assuming the silicon-oxygen framework to be the same in the two glasses, and adjusting for the change in absolute numbers of silicon and oxygen atoms, if necessary, one may entirely eliminate the radial distribution about silicon and oxygen. (Peaks due to the distribution about lithium would be small, and only minor errors would be introduced by subtracting them. These errors might be accounted for by computing the expected areas of the peaks.) One may also extend the differential technique to liquids and solutions, or, in some cases, to gases. In effect, with appropriate variations of composition, radial distribution curves may be computed for the density of scattering matter about any desired atomic species.

<sup>1</sup> Warren, B. E., and Biscoe, J., "Fourier Analysis of X-ray Patterns of Soda-silica Glass," *Jour. Amer. Ceramic Soc.*, 21, 259-265 (1938).

<sup>2</sup> Biscoe, J., and Warren, B. E., "X-ray Diffraction Study of Soda-boric Oxide Glasses," *Ibid.*, 21, 287-293 (1938).

<sup>3</sup> Green, R. L., "X-ray Diffraction and Physical Properties of Potassium Borate Glasses," *Ibid.*, 25, 83-89 (1942).

<sup>4</sup> Warren, B. E., Krutter, H., and Morningstar, O., "Fourier Analysis of X-ray Patterns of Vitreous SiO<sub>2</sub> and Vitreous B<sub>2</sub>O<sub>3</sub>," *Jour. Amer. Ceram. Soc.*, 19, 202-206 (1936).

<sup>5</sup> Buerger, M. J., "A New Fourier Series Technique for Crystal Structure Determination," *Proc. Nat. Acad. Sci.*, 28, 281-285 (1942).

## A NEW FOURIER SERIES TECHNIQUE FOR CRYSTAL STRUCTURE DETERMINATION

BY M. J. BUEGER

MINERALOGICAL LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated May 21, 1942

1. *Introduction.*—In an accompanying paper,<sup>1</sup> Joseph S. Lukesh demonstrates that important information can be derived about interatomic distances in glasses provided use is made of data from two related glasses. Specifically, if one prepares a curve by subtracting from the ordinates of the radial distribution curve of a two-component glass the ordinates of the radial distribution curve of one of the two components, there results a differential curve. This new curve has maxima at abscissae corresponding to the interatomic distances of the additional metal atoms in the two-component glass.

When the writer first learned of this discovery, it occurred to him that a corresponding difference function might be set up for two crystals of related structure and that this function would be capable of providing the interatomic vectors of the atom contained in one structure but absent in the other. A more careful examination of the situation, however, proves that an appropriate extension of the difference idea to Patterson diagrams gives rise to a new diagram having rather remarkable properties.

2. *Properties of the Patterson Diagram.*—Patterson has shown<sup>2,3</sup> that the Fourier summation

$$A_{x,y} = \sum \sum |F_{h k 0}|^2 e^{2\pi i(hx + ky)} \quad (1)$$

has peaks at points corresponding to the ends of interatomic vectors in the projection of the crystal structure. If a crystal has  $n$  atoms per unit cell, its two-dimensional Patterson diagram contains  $n^2$  peaks. Of this total,  $n$  peaks correspond with the several vectors from each of the atoms to itself; these  $n$  peaks all appear coalesced into one composite peak at the origin. The  $n$  atoms can be listed in pairs in  $n(n - 1)$  different ways, and to each of these there corresponds an interatomic vector (for each two atoms  $A$  and  $B$  there are two vectors,  $\overrightarrow{AB}$  and  $\overrightarrow{BA}$ ). The Patterson diagram, therefore, contains  $n(n - 1)$  non-origin peaks. The fact that this

diagram contains so many peaks for only  $n$  atoms per unit cell complicates its interpretation, and this limits its usefulness to cases where  $n$  is small.

3. *Properties of the Difference Diagram.*—Consider two crystals having the same structures but having a different species of atom occupying the same position in the two crystals, for example, the compounds  $MABCD$  and  $NABCD$ , where  $M$  and  $N$  occupy the same positions in the two structures. For the sake of simplicity in explanation, suppose that the cell contains only one formula weight. Then the Patterson diagram of  $MABCD$  contains peaks corresponding with the ends of the vectors representing the pairs  $MA, MB, MC, MD; AB, AC, AD; BC, BD; CD$ ; plus peaks corresponding to the reverse directions; plus peaks at the origin representing the distances between atom pairs  $MM, AA, BB, CC, DD$ . In the same way, the Patterson diagram of  $NABCD$  contains peaks corresponding with the ends of vectors representing the pairs  $NA, NB, NC, ND; AB, AC, AD; BC, BD; CD$ ; plus peaks corresponding to the reverse vectors; plus peaks at the origin representing the distances between atoms pairs  $NN, AA, BB, CC, DD$ . If the heights of these two Patterson diagrams are subtracted at each point, then the peaks common to the two diagrams vanish, leaving only peaks corresponding to the ends of the vectors  $(M - N)A, (M - N)B, (M - N)C, (M - N)D$ , plus peaks corresponding to the reverse vectors, plus a peak  $(M - N) (M - N)$  at the origin. This is a rather remarkable result, for it means that the difference diagram for a cell containing  $n$  atoms has only  $n$  peaks (of which  $n - 1$  are non-origin peaks) plus the  $n - 1$  centrosymmetrical peaks corresponding to reverse vectors.

If the projection of the crystal structure of  $MABCD$  contains only inversion centers and if  $M$  occupies one of these, then the peaks of the difference diagram occupy the same positions as atoms of the crystal structure, i. e., the difference diagram is an undistorted picture of the centers of the atoms in the crystal structure. If the crystal structure contains no symmetry, then the peaks of the difference diagram are the same as the positions of the atoms in the structure plus the positions of the atoms in the centrosymmetrical replica of the structure.

More generally, the difference diagram is related to the projected crystal structure in the following way: To derive the peaks of the difference diagram from the plane pattern, take the several symmetrically equivalent atoms  $(M - N)$  and their environs from their positions in their plane group and place these clusters at the center of the point group isomorphous with the plane group in such a way that clusters retain their mutual orientations and so that all atoms  $(M - N)$  fall at the origin of the point group. If the latter contains no inversion center, this must be introduced. The resulting atom positions are the same as the peaks in the difference diagram.

To interpret a difference diagram, the point group and plane group of the projected pattern must be known, and the problem resolves itself into one



of finding positions for the atoms on the plane group, knowing the corresponding positions on the point group. The chief obstacle to this is that the point group diagram is a composite picture of the clusters surrounding the  $(M - N)$  atoms in their correct position in the plane group, and question arises as to which set of atoms in the point group is this cluster. Given the several atoms of the composite cluster, there are, in general, several ways of combining them to give clusters about  $(M - N)$  in the plane group. One of these is the correct cluster, but all possibilities are consistent with the original data.

In general, then, it is less easy to pass from the difference diagram to the plane pattern projection of a crystal. Nevertheless, this problem is much easier than deriving the projected crystal pattern from the Patterson diagram, because for a unit cell containing  $n$  atoms, there are  $n(n - 1)$  peaks in the Patterson diagram, but only  $2(n - 1)$  peaks, at most, in the difference diagram. In many cases, the additional peaks in the Patterson diagram may be of aid in deciding between possible projected crystal structures predicted by the difference diagram, especially since the additional peaks are known to be of the type  $AB, AC, AD, BC$ , etc.

These principles may be extended to three-dimensional difference diagrams. Furthermore, a difference diagram made by subtracting the heights of Harker<sup>4</sup> diagrams has the interesting property of showing only peaks corresponding to interatomic distances of symmetry-related substitution atoms  $(M - N)$ . This diagram is, therefore, of aid in locating the substitution atoms  $(M - N)$ , about which the remainder of the atoms are clustered.

4. *Practical Derivation of Difference Diagrams.*—The difference diagram has been described as a diagram whose height at any given point is equal to the difference at the same point of the two Patterson diagrams for two crystals  $MABCD$  and  $NABCD$ , i. e.,

$$\Delta_{x,y; M-N} = A_{x,y; M} - A_{x,y; N} \quad (2)$$

$$= \sum \sum |F_{hk0; MABCD}|^2 e^{2\pi i(hx + ky)} - \sum \sum |F_{hk0; NABCD}|^2 e^{2\pi i(hx + ky)}. \quad (3)$$

By expansion, this proves to be

$$\Delta_{x,y; M-N} = \sum \sum (|F_{MABCD}|^2 - |F_{NABCD}|^2)_{hk0} e^{2\pi i(hx + ky)}. \quad (4)$$

It follows from this that, if one makes x-ray diffraction records for two crystals  $MABCD$  and  $NABCD$ , the value required for each term in the parentheses of (4) is simply the difference between measured  $F^2$  values for the same reflection for the two crystals. This makes it easy to carry out the synthesis directly without necessarily plotting first the Patterson diagrams of the two crystals. It also makes it possible to carry out the synthesis



optically<sup>5,6</sup> by simply making each hole in the appropriate grating with an area proportional to the difference of the  $F^2$ 's of the corresponding reflections of the two crystals. Optically negative holes<sup>5,6</sup> may be necessary in certain instances.

As a practical matter, the difference diagram need not be constructed from data obtained from two crystals having the substitute atom completely different. The two crystals may be somewhat differing members of a solid solution series in which the character of the solid solution is that of a single substitution. The origin is then the substituted atom.

5. *Application to Determining the Character of Solid Solutions.*—Solid solutions are structurally one of three types: Addition, subtraction or substitution or a combination of these types. Each of these may be formally regarded as substitution of  $N$  for  $M$ , where  $M$  and  $N$  are either real atoms or where one is a void. If a difference diagram is constructed using two members of a solid solution series as experimental crystals, the result is a difference diagram referred to the position of the substituted atom in the solid solution. From an analysis of this diagram, therefore, the environment of the locus of solid solution can be determined; and, if the general structural scheme is known, the type of solid solution structure can be determined.

6. *The Difference Diagrams for More Than One Substitution.*—In the preceding sections, properties of the difference diagram for the case of a single substitution have been discussed. If there are more substitutions, the number of peaks per unit cell are given by the middle column of the following table:

	1 SUBSTITUTION PER CELL OF $n$ ATOMS	$s$ SUBSTITUTIONS PER CELL OF $n$ ATOMS	$n$ SUBSTITUTIONS PER CELL OF $n$ ATOMS
Total number of peaks per cell, including the several unre- solved at the origin	$2n - 1$	$s(2n - s)$	$n^2$
Number of non-origin peaks	$2(n - 1)$	$s(2n - s - 1)$	$n(n - 1)$
Number of centrosymmetrical pairs of non-origin peaks	$n - 1$	$\frac{s}{2}(2n - s - 1)$	$\frac{n}{2}(n - 1)$

If there are  $n$  substitutions, the number of peaks per cell becomes identical with the number in a Patterson diagram.

<sup>1</sup> Lukesh, Joseph S., "Distribution of Metallic Atoms in Two-Component Glasses," *Proc. Nat. Acad. Sci.*, **28**, 277-281 (1942).

<sup>2</sup> Patterson, A. L., "A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals," *Phys. Rev.*, **46**, 372-378 (1934).

<sup>3</sup> Patterson, A. L., "A Direct Method for the Determination of the Components of Interatomic Distances in Crystals," *Zeit. Krist.*, (A) **90**, 517-542 (1935).

<sup>4</sup> Harker, David, "The Application of the Three-Dimensional Patterson Method and the Crystal Structure of Proustite,  $\text{Ag}_3\text{AsS}_3$ , and Pyrargyrite,  $\text{Ag}_3\text{SbS}_3$ ," *Jour. Chem. Phys.*, **4**, 381-390 (1936).

<sup>4</sup> Buerger, M. J., "The Photography of Interatomic Distance Vectors and of Crystal Patterns," *Proc. Nat. Acad. Sci.*, 25, 383-388 (1939).

<sup>5</sup> Buerger, M. J., "Optically Reciprocal Gratings and Their Application to Syntheses of Fourier Series," *Ibid.*, 27, 117-129 (1941).

## SYNTHESIS OF VITAMINS BY INTESTINAL BACTERIA

BY PAUL R. BURKHOLDER AND ILDA McVEIGH

OSBORN BOTANICAL LABORATORY, YALE UNIVERSITY

Communicated May 23, 1942

Recent studies on growth factors in microorganisms have contributed much to our general knowledge of nutrition and have provided a basis for devising quantitative microbiological assays for vitamins occurring in both plant and animal materials. Deficiencies of special growth factors for yeasts, molds and bacteria have received much attention recently, but the synthesis of vitamins by these non-green plants, of equally great significance, has not been extensively investigated.

The extent to which bacteria living normally in the alimentary tract of an animal may synthesize growth factors and contribute directly to the vitamin requirements of the animal constitutes a problem of some importance. The present paper represents an attempt to determine the approximate amounts of certain B vitamins produced by species of intestinal bacteria grown as pure cultures in a chemically defined medium.

Synthesis of vitamins has already been reported for a considerable number of bacteria. Snell and Strong<sup>1</sup> demonstrated synthesis of riboflavin by lactic acid bacteria. Silverman and Werkman<sup>2</sup> showed that certain propionic acid bacteria make thiamine or its intermediates. Some strains of dysentery bacilli<sup>3</sup> are able to form thiamine, and also Coenzyme I or II, riboflavin and perhaps biotin. It has been reported that a strain of diphtherial organisms<sup>4</sup> can make thiamine, Coenzyme I or II and riboflavin.

The evidence obtained by several investigators indicates that bacteria normally living in the rumina of herbivores, such as sheep and cattle, produce considerable quantities of vitamins. Thus it has been found that the common *Bacillus vulgatus* living in the intestines of herbivores is capable of synthesizing thiamine.<sup>5</sup> Recently it has been shown<sup>6</sup> that considerable amounts of riboflavin, pyridoxine and the antihemorrhagic vitamin are formed in the rumina of sheep and cows fed on diets low in these vitamins, and the source of the vitamins is assumed to be commensal microorganisms. Almquist, *et al.*,<sup>7</sup> demonstrated that common bacteria, such as *Bacillus subtilis* and *Escherichia coli*, can synthesize vitamin K. The phenomenon of refection which gives protection against certain vitamin deficiencies in labo-

ratory animals indicates that the normal intestinal flora of mammals may synthesize appreciable amounts of vitamins.

The intestinal bacteria employed in the present study were kindly supplied by Dr. George Valley, Department of Bacteriology, Yale University. The organisms, *Escherichia coli*, *Proteus vulgaris*, *Bacterium aerogenes*, *Alcaligenes fecalis*, *Bacillus mesentericus* and *B. vulgatus*, were grown as stock cultures on Difco nutrient agar. A special liquid medium prepared for the studies of growth factor production was made up as follows:  $K_2HPO_4$ , 1.0 gm.;  $MgSO_4 \cdot 7H_2O$ , 0.1 gm.; NaCl, 5.0 gm.;  $CaCl_2$ , 0.005 gm.; glucose, 10.0 gm.; recrystallized asparagine, 2.6 gm.; 1-tryptophane, 0.1 gm.; 1-cystine, 0.05 gm.; redistilled water, 1 liter. Small measured amounts of the following trace elements were added: Fe, Mn, B, Zn, Cu

TABLE 1

VITAMIN CONTENT OF BACTERIAL CULTURES GROWN FOR 48 HOURS AT 36°C. AND INOCULATED MEDIUM KEPT AT -2°C.

SPECIES OF BACTERIA	BIOTIN		RIBOFLAVIN		THIAMINE		NICOTINIC ACID	
	MILLI-GAMMA PER ML. OF MEDIUM	MILLI-GAMMA PER ML. OF CULTURE	GAMMA PER ML. OF MEDIUM	GAMMA PER ML. OF CULTURE	GAMMA PER ML. OF MEDIUM	GAMMA PER ML. OF CULTURE	GAMMA PER ML. OF MEDIUM	GAMMA PER ML. OF CULTURE
<i>Escherichia coli</i>	0	1.050	0	0.048	0.023	0.075	0.004	0.028
<i>Proteus vulgaris</i>	0	2.385	0.001	0.044	0.023	0.104	.....	Required
<i>Bacterium aerogenes</i>	0.015	2.370	0.004	0.140	0.005	0.148	0.005	0.300
<i>Alcaligenes fecalis</i>	0.028	0.446	0	0.067	0.038	0.146	0	0.066
<i>Bacillus mesentericus</i>	.....	Required	0	0.023	0.015	0.103	0	0.339
<i>Bacillus vulgatus</i>	0.028	1.365	0	0.136	0.031	0.150	0	1.181

and Mo. The medium was adjusted to pH 6.8 and sterilized by autoclaving at 15 lb. for 15 minutes. The inoculum was prepared by transferring a small amount of the organisms growing on agar to 5 ml. of physiological salt solution in test tubes. With a sterile pipette 0.1 ml. of the saline suspension was inoculated into 20 ml. of sterile culture liquid contained in small Erlenmeyer flasks. One set of inoculated flasks was kept as a control with growth inhibited in a room at -2°C., while the other was maintained in an incubator at +36°C. The period of growth was 48 hours unless stated otherwise.

*Proteus vulgaris* appears unable to synthesize nicotinic acid and *B. mesentericus* is deficient in biotin. It was found necessary, therefore, to add the deficient vitamin to the basal medium for these species in order to obtain growth so that tests could be made for the other vitamins which might be synthesized.

The methods used in assaying for riboflavin, biotin and nicotinic acid involved the use of *Lactobacillus casei*  $\epsilon$ , *Saccharomyces cerevisiae* F. B. and *Lactobacillus arabinosus* as indicators in microbiological tests described by Williams and others.<sup>8</sup> Thiamine activity was tested by the *Phycomyces* assay method.<sup>9</sup> Growth of the bacteria to be tested was measured with a turbidimeter, and the fresh weight of the cells in each culture was calculated by reference to standard fresh weight turbidity graphs prepared for all the species studied. At the end of the growth period, all cultures were acidified with sufficient concentrated  $H_2SO_4$  to make the liquid approximately 1 N. The acidified cultures were autoclaved at 15 lb. pressure for 30 minutes to effect hydrolysis of the cells. The samples were cooled, brought to pH 5.0 with NaOH and diluted to standard volume. The amounts of the solutions to be used in making the tests were determined by preliminary trials, and

TABLE 2

SYNTHESIZED VITAMIN RESIDUES IN CULTURES GROWN FOR 48 HOURS AT 36°C. VALUES EXPRESSED AS GAMMA PER GRAM OF FRESH CELLS

SPECIES	BIOTIN	RIBOFLAVIN	THIAMINE	NICOTINIC ACID
<i>E. coli</i>	2.3	106	115	62
<i>P. vulgaris</i>	3.2	57	95	None ?
<i>B. aerogenes</i>	1.1	41	43	89
<i>A. fecalis</i>	0.5	78	132	77
<i>B. mesentericus</i>	None ?	14	53	204
<i>B. vulgatus</i>	0.8	82	72	709

appropriate aliquots were employed at two concentration levels for each vitamin assay so that growth of the indicator organism would fall within a suitable range of response.

Some of the results obtained with vitamin assays performed on six species of bacteria are shown in table 1. Each value in the table represents the average of four or six determinations. The whole series of assays were repeated at different times on different cultures. In actual practice the test organisms gave satisfactorily consistent responses both to varied amounts of synthetic growth factors and to additions of bacterial extracts.

As indicated in table 1, the cultures which had grown at 36°C. for 48 hours showed a higher content of the four vitamins per ml. of fluid than did the inoculated medium in which growth was inhibited by low temperature. The results are taken to mean that under the conditions of the experiment these species of bacteria synthesize B vitamins in greater amounts than are used in their metabolism, and the residues accumulate in the cultures. The biotin, nicotinic acid, riboflavin and thiamine accumulated by the growing organisms were calculated as gamma per gram of fresh bacterial cells. These data are shown in table 2. The quantity of biotin was much lower

than the other growth factors found in the cultures. It is generally known, of course, that biotin exhibits biological activity in exceedingly small amounts. What the influence of different cultural conditions might be upon production of vitamins by bacteria would be worth further study.

An important question is how much of the total growth factors produced may be liberated from the cells into their environment. The few tests which have been made on filtrates and whole cultures of *E. coli* and *B. aerogenes* indicate that from 1 to 15% of the total biotin and nicotinic acid present in 48-hour cultures may occur outside the cells, while somewhat larger portions of the riboflavin and thiamine appear to be leached from the bacteria. In cultures of *B. aerogenes* which had grown for 111 hours, about 30%

of the biotin, thiamine and riboflavin and 40% of the nicotinic acid produced by the bacteria were found in the filtrate. Age of the cultures and other factors appear to be important in determining the distribution of these water-soluble vitamins between the microorganisms and the medium in which they live.

A study was made concerning the vitamins occurring in cultures of *B. aerogenes* at different periods of time up to 111 hours after inoculation of the basal medium. The results of this study are shown in the accompanying figure 1. It appears that the bacteria synthesized comparatively large amounts of vitamins during the early stages of growth. At 14½ hours the

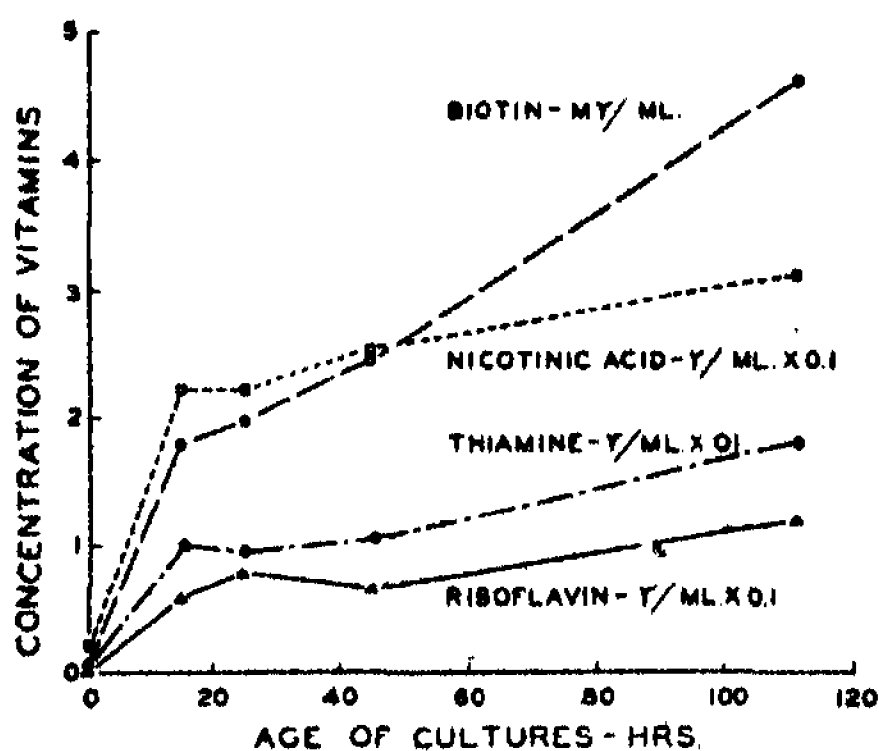


FIGURE 1

Concentration of vitamins in cultures of *B. aerogenes* grown in a chemically defined medium for different periods of time. Maximum values per ml. of culture suspension were as follows: biotin, 4.6 mγ; nicotinic acid, 0.31 γ; thiamine, 0.18 γ; riboflavin, 0.12 γ.

thiamine, nicotinic acid and riboflavin content of the cultures attained values almost as great as those reached subsequently up to 48 hours. The biotin content increased continuously throughout the entire period. From the standpoint of utilization of growth factors by the bacteria, it is significant that synthesis of the vitamins occurred early in the period of growth. Information of this kind should be valuable also in any attempt to estimate possible uses which may be made of the vitamins synthesized by bacteria.

In recent times the synthetic powers of microorganisms have been steadily assuming greater significance in relation to human welfare, as, for example, in the employment of bacteria and molds in the dairy and chemical industries and the use of yeast for the sake of its B vitamins as a supple-

ment in the diet of man. The extent to which microorganisms living normally in the alimentary tract of an animal may synthesize vitamins and contribute directly to the vitamin requirements of that animal constitutes a problem which has not yet received adequate attention. Demonstration of the synthesis of vitamins by intestinal bacteria, as presented in this report, should have considerable significance in connection with further investigations on the nutritional relationships existing among microorganisms, animals and man.

<sup>1</sup> Snell, E. E., and Strong, F. M., *Enzymologia*, 6, 186 (1939).

<sup>2</sup> Silverman, M., and Werkman, C. H., *Jour. Bact.*, 38, 25 (1939).

<sup>3</sup> Dorfman, A., Koser, S. A., Reames, H. R., Swingle, K. F., and Saunders, F., *Jour. Infect. Dis.*, 65, 163 (1939).

<sup>4</sup> Evans, W. C., Handley, W. R. C., and Happold, F. C., *Brit. Jour. Exp. Path.*, 20, 396 (1939).

<sup>5</sup> Scheunert, A., and Schieblich, M., *Biochem. Zeit.*, 139, 57 (1923).

<sup>6</sup> McElroy, L. W., and Gross, H., *Jour. Biol. Chem.*, 130, 437 (1930); *Jour. Nutrition*, 20, 527 (1939).

<sup>7</sup> Almquist, H. J., Pentler, C. F., and Mecchi, E., *Proc. Soc. Exp. Biol. Med.*, 38, 336 (1938).

<sup>8</sup> Williams, R. J., *University of Texas Publication No. 4137* (1941).

<sup>9</sup> Bonner, J., and Erickson, J., *Amer. Jour. Bot.*, 25, 685 (1938).

## THEORY OF THE EFFECTS OF LIGHT INTENSITY AND DURATION IN DETERMINING VISUAL RESPONSES

BY N. R. BARTLETT AND G. E. HUDSON<sup>1</sup>

DEPARTMENTS OF PSYCHOLOGY AND OF PHYSICS, BROWN UNIVERSITY

Communicated June 15, 1942

In this paper are presented some applications to the phenomena of brightness discrimination and absolute threshold measurements of a theory based partly on the familiar differential equation proposed by Hecht<sup>2</sup> to account for some of the phenomena of the sensory process. A generalized form of this equation is

$$\frac{dx}{dt} = \sum_{i=1}^N \left[ k_{1i} I (a_0 - x)^{m_i} - k_{2i} x^{n_i} \right],$$

where, for vision,  $I$  represents the intensity of the exciting light,  $x$  the concentration of photoproducts broken down from the original concentration  $a_0$  of the light-sensitive substance,  $t$  represents time,  $m_i$  and  $n_i$  are integral exponents indicating the order of the reactions and  $k_{1i}$  and  $k_{2i}$  are dimen-

sional constants characteristic of the individual eye. For simplicity in the practical applications we consider here only the bimolecular form of the equation, for which  $N = 1$ ,  $m_1 = n_1 = 2$ , and  $k_{11} = k_1$ ,  $k_{21} = k_2$ . It is to be noted, however, that the theory can be developed in an analogous manner for the more general equation in which both bimolecular and monomolecular reactions are included.

We assume that the threshold responses with which we are dealing occur when the concentration  $x$  changes by a constant amount  $\Delta x$ . To calculate this quantity it is necessary to find the solution  $x(t)$ , of the differential equation, appropriate to the given experimental conditions. In order to carry this out, we further assume that, whatever the variation of  $I$  with  $t$ , the solution,  $x(t)$ , is to be a continuous function of  $t$ . In case  $I$  is constant in some interval  $t_1 < t < t_2$ , then the general solution in this interval is

$$x = -\frac{1}{a} \frac{du}{dt} \frac{1}{u} = -\frac{1}{a} \left( \frac{u_1 U_1 e^{u_1 t} + u_2 U_2 e^{u_2 t}}{U_1 e^{u_1 t} + U_2 e^{u_2 t}} \right), \quad (1)$$

where  $u_1$ ,  $u_2$  and  $a$  are constants involving  $I$  and the fundamental constants appearing in the differential equation.  $U_1$  and  $U_2$  are arbitrary constants. Our investigations are restricted to those conditions where coupling effects between adjacent regions of the retina, such as inhibition and facilitation, may be neglected. Further we require that the angular areas of the retina excited by the light are small enough to be essentially homogeneous and are large enough that the response is an average of a large number of single cell (or nerve) responses.

The method used to derive the results presented here is applicable to a large class of problems. The experimental conditions defining this class are those in which the intensity  $I$  is constant in each of a finite or "infinite" number of time intervals,  $t_{i-1} < t < t_i$ . That is,  $I = I_i = \text{constant}$  in the  $i$ th interval.

The simplest case to be treated here is that of absolute threshold with a dark-adapted eye. After this we consider the more general condition in which there is a change, at the instant  $t = 0$ , from the prevailing intensity  $I'$  to the intensity  $I''$ . This case includes, as special cases, the preceding one and the final one to be considered. In the last case we derive the minimum time,  $\Delta t$ , for the prevailing brightness, of intensity  $I'$ , to be obliterated when a "dark" flash,  $I'' = 0$ , is detected.

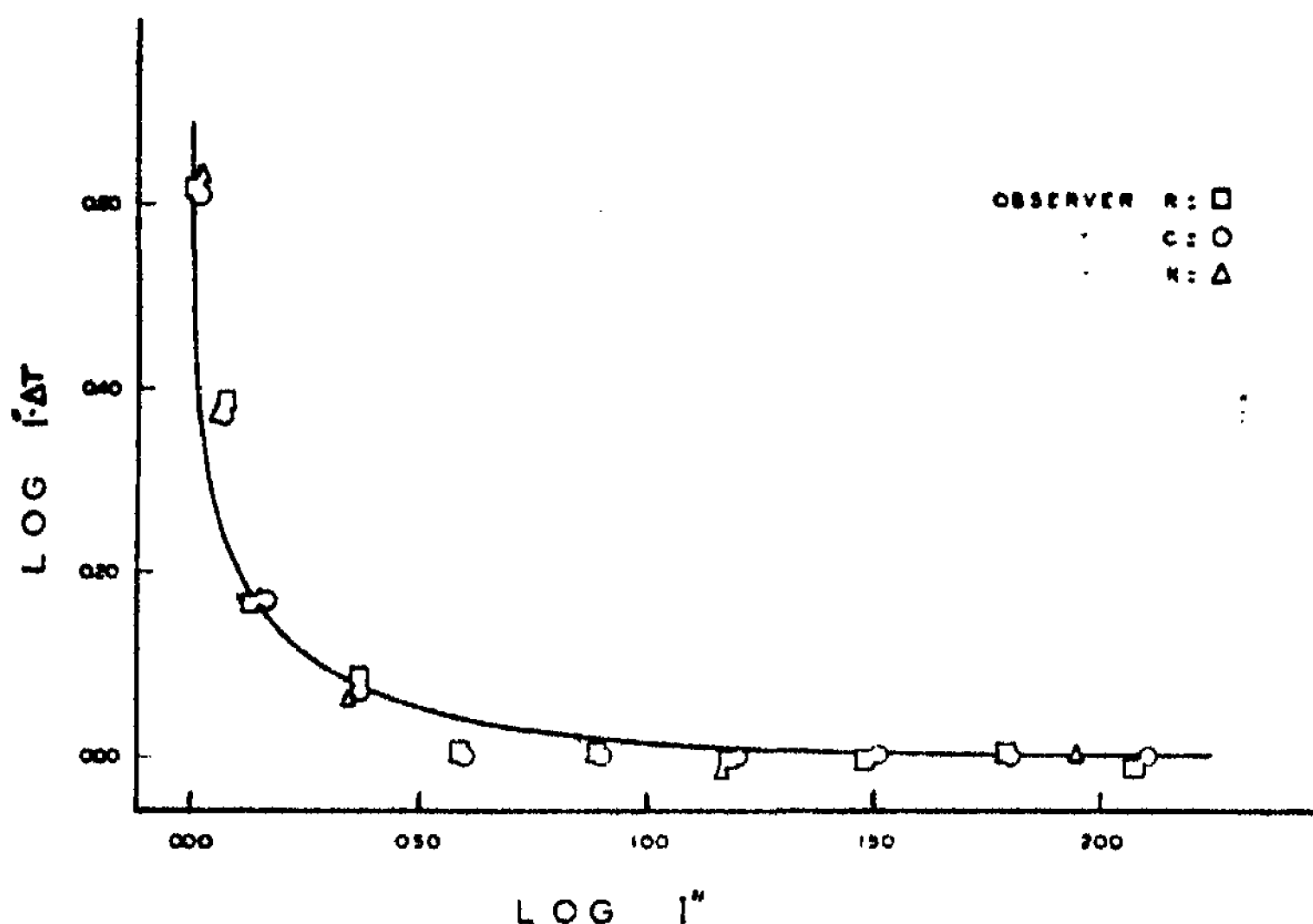
*I. Absolute Threshold.*—The experimental conditions require that, for a long while previous to the time  $t = 0$ ,  $I = I' = 0$ . At the instant  $t = 0$ ,  $I$  becomes equal to  $I''$  and remains constant until the instant  $\Delta t$ , when  $I$  again becomes equal to zero.  $\Delta t$  is the minimum exposure time required for the detection of the flash. The continuity condition requires that at  $t = 0$ ,  $u [= U_0 e^{-\int a x dt}]$  and  $(1/a)(du/dt)$  are continuous.



Applying these conditions, we find after some algebraic manipulation an expression for  $\Delta x$  in terms of  $I''$  and  $\Delta t$ . Since we are considering only bi-molecular reactions, solution for  $\Delta t$  of this expression yields

$$\Delta t = \frac{B}{2\sqrt{I''}} \log \frac{G\sqrt{I''} + 1}{G\sqrt{I''} - 1}, \quad (2)$$

where  $B$  represents the constant  $1/(a_0\sqrt{k_1k_2})$  and  $G$  the constant  $[(a_0/\Delta x) - 1]\sqrt{k_1/k_2}$ . If we suppose  $\sqrt{I''} = (1/G)$ , then  $\Delta t$  becomes infinite; or if we suppose  $G\sqrt{I''} \gg 1$ , then  $I''\Delta t \rightarrow B/G$ . Hence we are led to the empirical laws of  $I = C$  and  $I\Delta t = C$  for long and short durations, respectively.



The relation between  $\log I''\Delta t$  and  $\log I''$  for absolute threshold, using a twenty-minute foveal area. The curve is described by equation (2). These data were obtained from Karn.<sup>3</sup>

By a simple transformation of equation (2) it may be seen that the constants  $B$  and  $B/G$  can be read directly from a plot of  $\log I''\Delta t$  vs.  $\log I''$ , obtained from the data of any given observer. It is seen that the curve for an observer is translated parallel to itself when the asymptotic constants for that observer are changed. Since the shape of the theoretical curve does not depend on any of these constants characterizing the individual, the curves for all observers should coincide when the experimental data are so translated by the appropriate addition of these constants. The congruence is obvious in the accompanying figure, showing Karn's data for three observers.<sup>3</sup> Karn obtained these data using a circular area subtending a visual angle of twenty minutes in diameter. The agreement with the curve derived from equation (2) is satisfying. Furthermore, by extra-



polating from Karn's experimental data, we may expect that the agreement with a somewhat larger area, e.g., sixty minutes, would be excellent.

*II. Brightness and Darkness Discrimination.*—The measurement of absolute thresholds is a special instance of the general case where some intensity  $I'$  has prevailed over the particular retinal area for a long time antecedent to the instant  $t = 0$ . In the preceding case, absolute threshold, this prevailing intensity is zero. If a flash  $\Delta I$  is added (or subtracted) to  $I'$  at the instant  $t = 0$  to render a new intensity  $I'' = I' + \Delta I$  for a duration  $\Delta t$ , we find

$$\Delta t = \frac{B}{2\sqrt{I''}} \log \left\{ \frac{1 + H \frac{(1 + R\sqrt{I'}) (1 - R\sqrt{I''})}{\sqrt{I'} + \sqrt{I''}}}{1 + H \frac{(1 + R\sqrt{I'}) (1 + R\sqrt{I''})}{\sqrt{I'} - \sqrt{I''}}} \right\}, \quad (3)$$

where  $H = |\Delta x/a_0| \sqrt{k_2/k_1}$ , if  $I'' > I'$ , and  $H = -|\Delta x/a_0| \sqrt{k_2/k_1} = -N$ , if  $I'' < I'$ , and  $R = \sqrt{k_1/k_2}$ . When  $I' = 0$ , this equation reduces at once to (2).

*III. "Inverse" Absolute Threshold.*—If now, in equation (3), we suppose  $I'' \rightarrow 0$  throughout the interval  $0 < t < \Delta t$ , then  $H = -N < 0$ , and we have in the limit

$$\Delta t = \frac{B\hat{N}}{\sqrt{I'}} \frac{(1 + R\sqrt{I'})^2}{\sqrt{I'}(1 - NR) - N}. \quad (4)$$

As  $I' \rightarrow \infty$ ,  $\Delta t \rightarrow R^2(B/G)$ , and as  $\sqrt{I'} \rightarrow N/(1 - NR) = 1/G$ ,  $\Delta t \rightarrow \infty$ . Equation (4) is the theoretical relation describing the data to be expected from measurement of the minimum time of obliteration of a prevailing intensity for detection of a "dark" flash. We have chosen to call this time "inverse" absolute threshold. It is hoped that experimental tests of this relation will soon be forthcoming.

The constants  $G$ ,  $B/G$  and  $R$  may be determined empirically from data obtained in experiments performed under the conditions described in I and III. The value of  $N$  (or  $H$ ), for the particular observer characterized by these constants, follows immediately from the relation  $H = 1/(G + R)$ . Data from brightness discrimination experiments for this observer may then be examined to determine whether they are adequately described by the theoretical curve computed from equation (3), using these values of the constants. If the comparison is satisfactory, it may be considered as an experimental check of the values of the constants and of the theory outlined here.

<sup>1</sup> Bartlett, N. R., and Hudson, G. E. (manuscript in preparation).

<sup>2</sup> Hecht, S., *Physiol. Rev.*, 17, 239-289 (1937).

<sup>3</sup> Karn, H. W., *Jour. Gen. Psychol.*, 14, 360-369 (1936).

## STATIONARY POINTS OF TRANSFORMATION GROUPS

BY P. A. SMITH

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Communicated June 2, 1942

Let  $(G, S)$  be a continuous realization of the topological group  $G$  by topological transformations of the space  $S$  into itself. The word "realization" is to be understood in its broadest sense: The transformation group  $(G, S)$  is a homomorphic, but not necessarily isomorphic, image of  $G$ . Points of  $S$  which are fixed under all the transformations in  $(G, S)$  will be called stationary and their totality denoted by  $\sigma(G, S)$ . We shall examine the structure of  $\sigma$ .

*Definition 1.* A compact finite-dimensional space  $S$  will be called an  $(n, p)$ -sphere ( $n \geq 1$ ) if (1)  $S$  has the same homology groups, coefficients mod  $p$ , as an  $n$ -sphere; (2) no proper subset of  $S$  satisfies (1). A  $(0, p)$ -sphere is a pair of points; the empty set is a  $(-1, p)$ -sphere. It can be shown that the dimension of a locally euclidean  $(n, p)$ -sphere is  $n$ .

Let  $G$  be a topological group, and let  $n, p$  be non-negative integers. If for every  $(n, p)$ -sphere  $S$  and realization  $(G, S)$ ,  $\sigma(G, S)$  is an  $(m, p)$ -sphere ( $-1 \leq m \leq n$ ), we shall say that *proposition*  $[G, n, p]$  is true. If every realization  $(G, E_n)$  ( $E_n$  = euclidean  $n$ -space) admits at least one stationary point, we say that *proposition*  $[G, n]$  is true.

*I. Proposition*  $[G, n, p]$  holds for prime  $p$  and arbitrary  $n$  if  $G$  is a finite abelian group whose order is a power of  $p$ .

If  $G$  is cyclic, I follows from known theorems concerning periodic transformations.<sup>1</sup> Starting with this, the proof of I is similar to that of

*II. Proposition*  $[G, n, p]$  holds for prime  $p$ , arbitrary  $n$  and connected compact abelian Lie group  $G$ .

*Proof.* Let  $I_\alpha$  or  $I(\alpha)$  denote the fixed-point set of  $\alpha$ —that is, of the transformation  $x \rightarrow \alpha \cdot x$  ( $\alpha \in G, x \in S$ ). The subgroup  $G^p = \{\alpha\}$  of elements of  $G$  whose orders are powers of  $p$  is everywhere dense in  $G$ . By I,  $I_\alpha$  is an  $(h, p)$ -sphere,  $h = h(\alpha)$ . Since  $h = n$  would imply  $I_\alpha = S$  which, if true for every  $\alpha$ , implies  $\sigma(G, S) = S$ , we may suppose that for some  $\alpha$ ,  $h$  is smaller than  $n$ . Since  $G$  is abelian,  $a \cdot I_\alpha = I_\alpha$  for every  $a \in G$ . Thus there is induced in  $I_\alpha$  a realization  $(G, I_\alpha)$  and evidently  $\sigma(G, S) = \sigma(G, I_\alpha)$ . The theorem now follows by induction on  $n$ .

*III. Proposition*  $[G, n]$  holds for arbitrary  $n$  and compact connected abelian Lie group  $G$ .

For, on adjoining a point  $P_\infty$  to  $E_n$ , we obtain a realization  $(G, S)$  where  $S$  is an ordinary  $n$ -sphere. Since  $P_\infty$  is stationary,  $\sigma(G, S)$  is non-empty, hence contains at least two points. Hence  $\sigma(G, E_n)$  contains at least one point.

We now consider realizations of non-abelian groups. Whether proposition  $[G, n, p]$  holds for compact connected Lie groups  $G$ , arbitrary  $n$  and some  $p$ , is at present a matter of conjecture. We answer the question in various special cases.

**Definition 2.** We shall call  $(G, S)$  *normal* if to each element  $a$  of  $G$  we can associate an element  $\alpha$  of order 2 such that  $I_a = I_\alpha$ .

**Example.** Suppose that  $G$  is a compact connected Lie group and  $A$  is a maximal connected abelian subgroup of  $G$ . If the induced realization  $(A, S)$  is normal then so is  $(G, S)$ . This follows from the fact that two  $A$ 's are transformable one into the other (Cartan, Weil<sup>2</sup>).

**IV.** *Within the class of normal realizations, proposition  $[G, n, 2]$  holds for every compact Lie group  $G$  and  $n \geq 0$ .*

**Proof.** Let "sphere" mean " $(m, 2)$ -sphere." By normality and theorem I, every  $I_a$  is a sphere. Let  $\{\alpha\}$  be the elements of order 2 in  $G$ . By induction on  $k$ , we see that intersections of  $k$  sets  $I(\alpha)$ —hence by normality, intersections of  $k$  sets  $I(a)$ —are spheres. For let  $I = \bigcap_{i=1}^k I(\alpha_i)$ . Since  $\alpha_k \cdot I(\alpha_k \alpha_i) = I(\alpha_k^{-1}(\alpha_k \alpha_i)\alpha_k) = I(\alpha_i \alpha_k) = I(\alpha_k \alpha_i)^{-1} = I(\alpha_k \alpha_i)$ ,  $\alpha_k$  induces a transformation in the set  $Q = \bigcap_{i=1}^{k-1} I(\alpha_k \alpha_i)$ , which is a sphere by normality and the hypothesis of induction. The fixed-point sphere is  $I(\alpha_k) \cap Q$  which, one can easily see, is precisely  $I$ . The theorem now follows readily from the fact that  $\sigma = \bigcap I_a, a \in G$ .

**Definition 3.** The *locally euclidean case* is that in which the space under transformation is locally euclidean. If in addition the functions defining  $(G, S)$  are analytic relative to families of analytically connected local coordinate systems in  $G$  and  $S$ , we call  $(G, S)$  *analytic*.

**V.** *In the locally euclidean case proposition  $[G, 3, 2]$  holds for every compact connected Lie group  $G$ .*

**Proof.** Let  $\xi$  be a point of  $\sigma$ . Since  $G$  is connected, orientation in the neighborhood of  $\xi$  cannot be reversed by any element  $a$  of  $G$ . Hence for  $\alpha$  of order 2,  $I_\alpha$  is either the  $(3, 2)$ -sphere  $S$  or a simple closed curve. As in the proof of IX below, we may assume that the homomorphism  $G \rightarrow (G, S)$  is one-one. Hence  $I_\alpha \neq S$ . Let  $A$  be a maximal connected abelian subgroup of  $G$ . There is induced in the simple closed curve  $I_\alpha (\alpha \in A)$  a realization  $(A, I_\alpha)$  (cf. proof of II). If  $\beta$  is an element in  $A$  of any finite order, its fixed-point set cannot be empty, otherwise  $\sigma(G, S) = 0$ ; nor can it be a pair of points, otherwise  $\beta$  reverses orientation in  $I_\alpha$ . The only remaining possibility is that  $I_\beta = I_\alpha$ . Since the elements  $\beta$  are everywhere dense in  $A$ , we have  $I_b = I_\alpha$  for every  $b \in A - \{e\}$ . Hence  $(A, S)$  is normal and hence  $(G, S)$  is normal (see example above).

**VI. (Corollary).** *Proposition  $[G, 3]$  holds for every connected compact Lie group  $G$ .*

Suppose  $(G, S)$  is analytic. Let  $n = \dim S$  at  $\xi$ ,  $\xi$  a point of  $\sigma(G, S)$ . In the space  $D_n$  of differentials at  $\xi$ ,  $(G, S)$  induces a representation  $(G, D_n)$  of

$G$  by real linear homogeneous transformations. Let  $J$  denote fixed point sets in  $D_n$ . It can be shown that  $\dim J_a = \dim I_a$  in the neighborhood of  $\xi$ .

*Definition 4.* An analytic  $(G, S)$  is *normal at  $\xi$*  ( $\xi \in \sigma$ ) if the induced  $(G, D_n)$  is normal.

VII. In the analytic case theorem IV holds if normality is replaced by normality at some point  $\xi$  of  $\sigma$ .

Let  $R_n$  denote the group of rotations in  $E_n$ .

VIII. In the analytic case, proposition  $[R_n, n, 2]$  holds for every  $n \geq 0$ .

In fact,  $(R_n, S_n)$  is normal at every point  $\xi$  of  $\sigma$ .

IX. In the analytic case, proposition  $[G, n, 2]$  holds for every compact connected Lie group  $G$  and  $n \leq 5$ .

Sketch of proof of  $[G, 5, 2]$ . Let  $G_0$  be the kernel of the natural homomorphism  $G \rightarrow (G, S)$  and let  $H = G/G_0$ . We define in a natural way an analytic realization  $(H, S)$  such that the correspondence  $H \rightarrow (H, S)$  is one-one and such that  $\sigma(H, S) = \sigma(G, S)$ . Let  $\xi$  be a point of  $\sigma(H, S)$ . The representation  $(H, D_\xi)$  induced at  $\xi$  has a decomposition

$$(H, D_\xi) = (H, D_h) + (H, D_\lambda) + (H, D_\mu) + \dots$$

where  $h + \lambda + \mu + \dots = 5$  and where  $(H, D_h)$  is the trivial representation of degree  $h$ , and the remaining representations are irreducible. From II, we may assume that  $G$  is non-abelian. Denoting by  $R_3'$  the covering group of  $R_3$ , an examination of the possible irreducible real representations of compact groups, of degrees  $\leq 5$  shows that the only non-trivial possibilities are, up to equivalences,<sup>3</sup>

1.  $H = R_3', (H, D_\xi) = (R_3', D_1) + (R_3', D_4)$
2.  $H = R_3, (H, D_\xi) = (R_3, D_2) + (R_3, D_3)$
3.  $H = R_4, (H, D_\xi) = (R_4, D_1) + (R_4, D_4)$
4.  $H = R_3, (H, D_\xi) = (R_3, D_5)$
5.  $H = R_5, (H, D_\xi) = (R_5, D_5)$

The representations  $(R_3', D_1)$ ,  $(R_3, D_2)$  and  $(R_4, D_1)$  are trivial. The remaining representations are irreducible; they exist and are unique. An examination of the weights of  $(R_3', D_4)$  shows that it is normal. Since  $(R_3, D_3)$ ,  $(R_4, D_4)$ ,  $(R_5, D_5)$  are also normal, it follows that  $(H, D_\xi)$  is normal in cases 1, 2, 3, 5, and IX follows from VII.

There remains only case 4; this realization is not normal. Take as generators of  $R_3$ , the infinitesimal rotations  $A_1, A_2, A_3$  about mutually orthogonal axes in ordinary space. Let  $\Gamma_i$  be the 1-parameter group generated by  $A_i$ . There is a unique element  $\alpha_i$  of order 2 in  $\Gamma_i$  and  $\{e, \alpha_1, \alpha_2, \alpha_3\}$  is a subgroup of  $R_3$ . Let  $a_i$  denote an arbitrary element in  $\Gamma_i - \{e, \alpha_i\}$ . We have  $\alpha_i a_j \alpha_i = a_j^{-1}$  ( $i \neq j$ ). Let  $\tau$  be an element of order 3 such that the

transformation  $a \rightarrow \tau a \tau^{-1}$  permutes the  $\alpha_i$  cyclically. It is easy to see that such a  $\tau$  exists.

In a suitable  $x_i$ -coördinate system in  $D_6$ , the infinitesimal elements corresponding to the  $A_i$  are given by<sup>4</sup>

$$2A_1^0 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ -4 & 0 & 2 & 0 & 0 \\ 0 & -6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad 2A_2^0 = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & -6 & 0 \\ 4 & 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \quad 2A_3^0 = \begin{pmatrix} 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

These matrices can be obtained from the well-known representations of the unitary unimodular group in two variables; that they actually constitute a representation of  $R_3$  can be verified by computing the commutators. Denoting fixed-point sets in  $D^6$  by  $J$ , an examination of the characteristic vectors of the  $A_i^0$  shows that (i) the sets  $J(\alpha_i)$  are 3-dimensional; (ii)  $J(\alpha_i)$  is independent of  $\alpha_i$  in  $\Gamma_i - \{\epsilon, \alpha_i\}$  and the three sets  $J(\alpha_i)$  are distinct lines.

The fixed-point sets in  $S_6$  are analytic loci. The sets  $I(\alpha_i)$  are 3-dimensional in the neighborhood of  $\xi$  and being fixed-point sets of transformations of period 2, they are (3, 2)-spheres and locally euclidean 3-dimensional throughout. Since  $\alpha_1 \cdot I(\alpha_2) = I(\alpha_1 \alpha_2 \alpha_1) = I(\alpha_3 \alpha_1) = I(\alpha_2)$  and since the fixed-point set of the transformation that  $\alpha_i$  thus induces in  $I(\alpha_j)$ ,  $i \neq j$ , is  $I(\alpha_i) \cap I(\alpha_j)$ , it follows that this last set, which is 2-dimensional in the neighborhood of  $\xi$ , is a (2, 2)-sphere and locally euclidean 2-dimensional throughout, hence is homeomorphic to an ordinary sphere. The relations  $I(\alpha_i) \cap I(\alpha_j) \subseteq I(\alpha_i \alpha_j) = I(\alpha_k)$  ( $i, j, k$  distinct) imply that the three spheres  $I(\alpha_i) \cap I(\alpha_j)$  are identical. Let them be denoted by  $\mu$ . The sets  $I(\alpha_i)$  are analytic (1, 2)-spheres, hence simple closed curves. Since  $\alpha_1 \cdot I(\alpha_2) = I(\alpha_1 \alpha_2 \alpha_1) = I(\alpha_2^{-1}) = I(\alpha_2)$ ,  $\alpha_1$  either leaves  $I(\alpha_2)$  pointwise fixed or induces in it a transformation of period 2. It is easy to show that only the first alternative can hold; hence  $I(\alpha_i) \subseteq \mu$ . Now by (ii),  $I(\alpha_1)$  and  $I(\alpha_2)$  intersect with distinct tangents at  $\xi$ . Since  $\mu$  is a sphere, there must be at least a second point  $\xi'$  of intersection. Since  $\xi'$  is stationary under the group generated by the elements  $\alpha_1$  and the elements  $\alpha_2$ , and since that group is  $R_3$ , we conclude that  $\sigma$  contains at least two points. Consider the element  $\tau$  defined above. Evidently  $\tau$  permutes the sets  $I(\alpha_i)$  cyclically, as well as the sets  $I(\alpha_i)$ . We conclude first that  $\tau \cdot \mu = \mu$  and secondly that  $\tau$  is not the identity over  $\mu$ . Hence  $\tau$  is of period 3 over  $\mu$  and admits exactly 2 fixed points in  $\mu$ . Hence  $\sigma$  contains at most two points in  $\mu$ . But  $\sigma \subseteq I(\alpha_1) \cap I(\alpha_2) = \mu$ . Hence  $\sigma = \{\xi, \xi'\} =$  a (0, 2)-sphere.

<sup>4</sup> Smith, P. A., "Transformations of Finite Period," *Ann. Math.*, 39, 137-164 (1938); "Transformations of Finite Period. II," *Ann. Math.*, 40, 690-711 (1939). See also Appendix B in Lefschetz, *Algebraic Topology* (Colloquium Publications of the Am. Math. Soc.) soon to appear.

<sup>2</sup> Weil, A., "Démonstration topologique d'un théorème fondamental de Cartan," *C. R.*, 200, 518-519 (1935).

<sup>3</sup> Cartan, E., "Les groupes projectifs continus réels qui ne laissent invariante aucune multiplicité plane," *Jour. de Math.*, 10, 149-186 (1914).

<sup>4</sup> Explicit formulae for this representation are not really essential to the proof.

### NOTE ON THE *t*-TEST

BY EDWIN B. WILSON AND JANE WORCESTER

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated June 15, 1942

For a normal universe with zero mean and standard deviation  $\sigma$ , the distribution of the absolute value of the mean of random samples of  $n' = n + 1$  and of the estimated value  $s = [\Sigma(x - m)^2/n]^{1/2}$  of  $\sigma$  is

$$\frac{2\sqrt{n'}}{\sqrt{2\pi}\sigma} e^{-n'm^2/2\sigma^2} dm \frac{2n^{n/2}}{2^{n/2}\Gamma(n/2)\sigma^n} e^{-ns^2/2\sigma^2} s^{n-1} ds \quad (1)$$

with limits 0 to  $\infty$  for both  $m$  and  $s$ . In terms of  $\sigma$  the test of significance for the mean  $m$  is that  $T = \sqrt{n'} m/\sigma$  be considered a normal unit variate (so that at the level  $P = 0.05$ ,  $T = 1.96$ ). If the value of  $\sigma$  is unknown, the ratio  $t = \sqrt{n'} m/s$  is set up,  $m$  is eliminated from (1) and an integration with respect to  $s$  leads to Student's distribution, doubled for the absolute value of  $t$ ; tables for the  $t$ -test give the values of  $t$  for certain levels of significance for values of  $n$  from 1 to 30.

We propose briefly to discuss the frequency distribution for  $T$  and  $t$  obtained from (1) by eliminating  $s$ , namely,

$$\frac{4n^{n/2}}{\sqrt{2\pi} 2^{n/2}\Gamma(n/2)} e^{-1/2(1 + n/t^2)T^2} \frac{T^n}{t^{n+1}} dTdt \quad (2)$$

for which the limits of both  $T$  and  $t$  are 0 and  $\infty$ . It is clear that although  $m$  and  $s$  are independent in (1), the variables  $T$  and  $t$  are not independent in (2). The frequency function vanishes when  $T = 0$  and when  $t = 0$ , except at  $T = t = 0$  where it is indeterminate, taking any positive value for the proper approach. The mode and mean of  $T$  as functions of  $t$  are<sup>1</sup>

$$T_{\text{mode}} = \frac{\sqrt{n}}{(1 + n/t^2)^{1/2}}, \quad T_{\text{mean}} = \frac{\sqrt{2}\Gamma(n/2 + 1)}{\Gamma(n/2 + 1/2)} \frac{1}{(1 + n/t^2)^{1/2}} \quad (3)$$

These equations are not rectilinear. The moment of order  $p$  in  $T$  and  $q$  in  $t$  for (2) about the origin is<sup>2</sup>

$$\nu_{pq} = \frac{2^{p/2} n^{q/2}}{\sqrt{\pi} \Gamma(n/2)} \Gamma\left(\frac{p+q+1}{2}\right) \Gamma\left(\frac{n-q}{2}\right). \quad (4)$$

The correlation coefficient between  $T$  and  $t$  is<sup>3</sup>

$$r = \frac{\sqrt{1 - 2/\pi}}{\sqrt{\frac{2\Gamma^2(n/2)}{(n-2)\Gamma^2(n/2 - 1/2)} - \frac{2}{\pi}}} \text{ or } r = \sqrt{\frac{n-2}{2} \frac{\Gamma(n/2 - 1/2)}{\Gamma(n/2)}} \quad (5)$$

according as  $T$  and  $t$  are limited to positive values or are allowed to range from  $-\infty$  to  $+\infty$ ; the correlation ratio of  $T$  about its regression curve is

$$\eta = \sqrt{\frac{\frac{\Gamma^2(n/2 + 1)}{(n/2 + 1/2)\Gamma^2(n/2 + 1/2)} - \frac{2}{\pi}}{1 - (2/\pi)}} \text{ or } \sqrt{\frac{\Gamma^2(n/2 + 1)}{(n/2 + 1/2)\Gamma^2(n/2 + 1/2)}} \quad 1$$

in the two respective cases.

The relative frequency for  $T$  when  $t$  is given is

$$\frac{2(1 + n/t^2)^{(n+1)/2}}{2^{(n+1)/2} \Gamma(n/2 + 1/2)} e^{-1/2(1 + n/t^2)T^2} T^n dT \quad (6)$$

which is a  $\chi^2$  distribution with  $\chi^2 = (1 + n/t^2)T^2$  and with  $n' = n + 1$  degrees of freedom. Thus one may find from  $\chi^2$  tables, if sufficiently detailed, what is the chance that  $T$  exceed a specified value when  $t$  is known from observation. For example, for the smallest sample ( $n = 1$ ,  $n' = 2$ ), the chance that  $T$  exceed 1.96 is 0.01 for  $t = 0.846$ , 0.05 for  $t = 1.34$ , 0.10 for  $t = 2.23$  and is only 0.155 for  $t = \infty$ , although  $t = 12.71$  is significant at the 0.05 level. For  $n = 2$  the tabular value  $t = 4.303$  corresponding to  $P = 0.05$  will ensure that  $T > 1.96$  in somewhat over 20% of the samples but no value of  $t$  is large enough to ensure it in 30% of the samples. For  $n = 4$  the tabular value 2.776 will ensure that  $T > 1.96$  in just over 30% of the samples. For  $n = 8$ , the tabular value 2.306 at  $P = 0.05$  will ensure that  $T > 1.96$  in something like 37% of the samples but no value of  $t$  can be so high as to ensure it in 95% of the samples. As the number in the sample increases indefinitely the value of  $t$  approaches 1.96 and the percentage of the samples in which  $T > 1.96$  approaches 50. For  $n = 29$  (a sample of 30) the value of  $t$  which will ensure that  $T > 1.96$  in 50% of the samples is 2.09, whereas the value of  $t$  corresponding to 0.05 is 2.05.

In figure 1 is represented the  $S$ - $T$ -plane for samples of 5 with  $T$  and  $S = s/\sigma$  as abscissa and ordinate. The lines  $T = 1.96$ ,  $t = 2.776$  are shown which divide the significant from the non-significant part of the plane at



$P = 0.05$ ; as are also the lines  $S = 1.67$  and  $S = 0.348$  beyond which on the upper and lower sides, respectively, the frequency function (1) has an integral value of 0.025. These four lines divide the plane into ten regions: In I-V  $T$  is not significant, in VI-X it is; in I, II, III, VI, VIII,  $t$  is not significant, in IV, V, X, IX, VII, it is; in II, IV, VIII, IX,  $s$  lies between the 2.5 and 97.5 percentiles of its distribution, in I, VI, VII, III, V, X, it lies outside those limits. The curved line in the diagram is the 0.05 contour line of the frequency function (1); for the two-dimensional normal distribution (Bravais) the contour line specified by any value of the frequency function leaves outside of it a total probability equal to that value, and in this

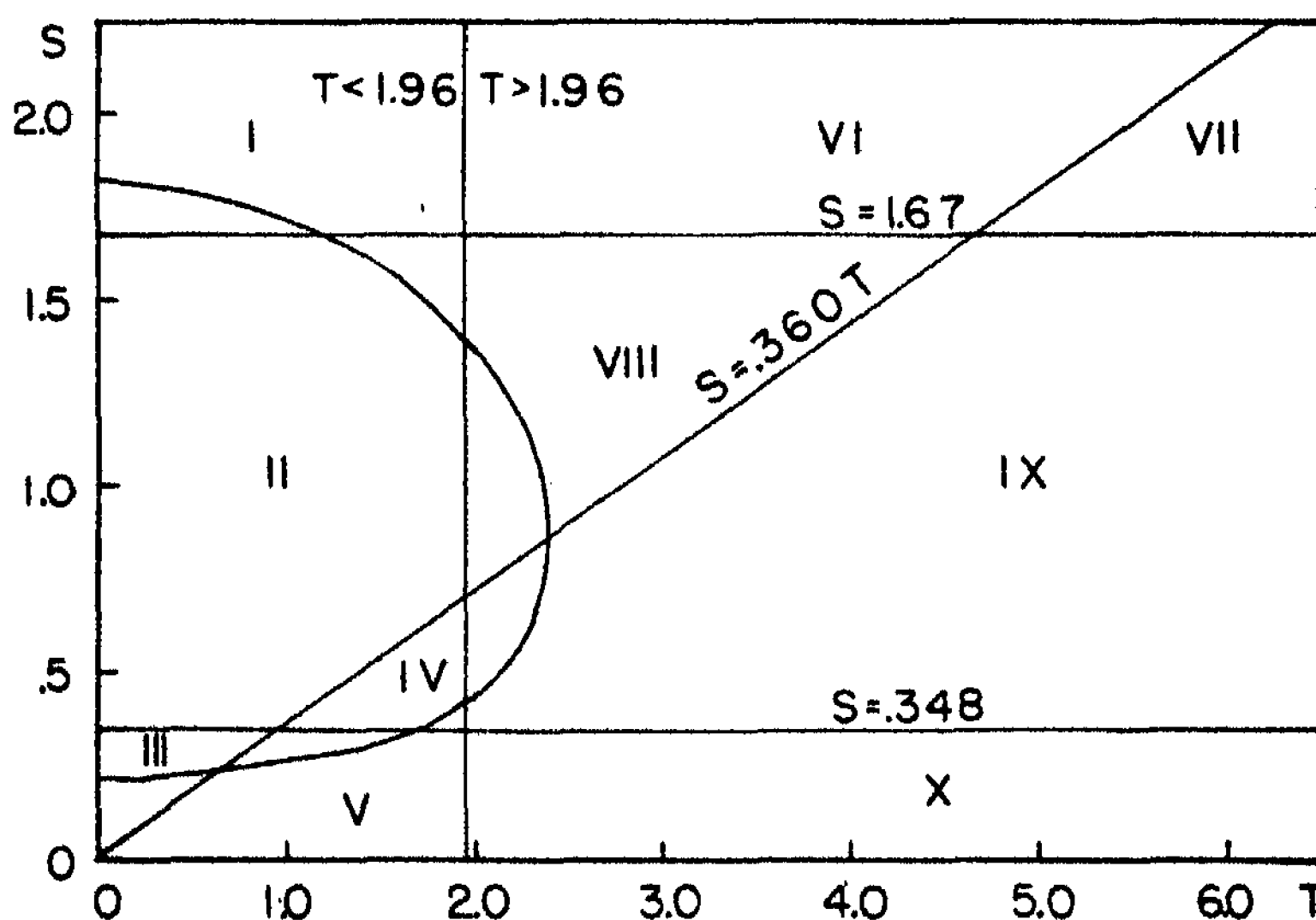


FIGURE 1

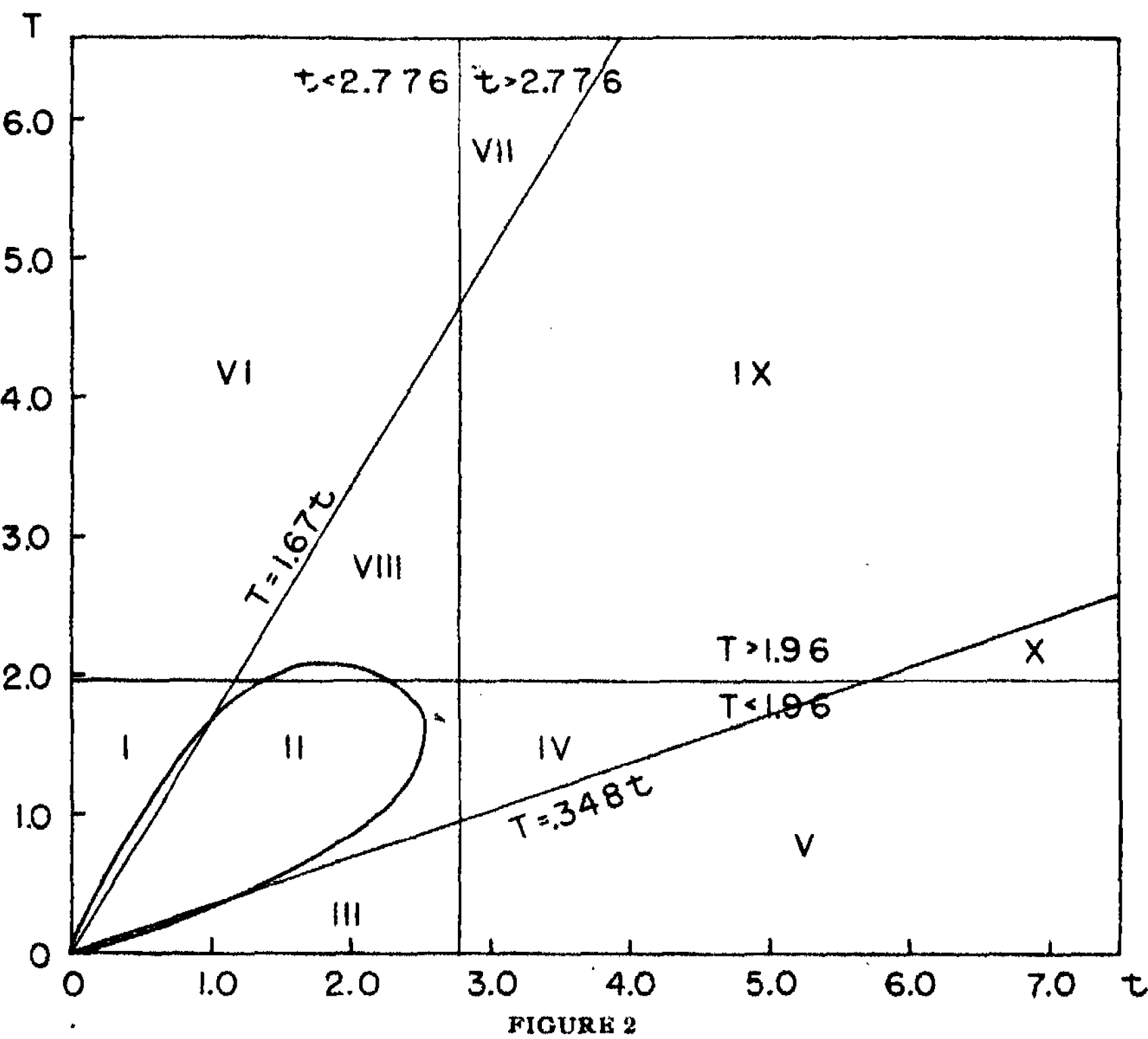
Regions of the  $S$ - $T$ -plane connected with significance of mean, standard deviation and the ratio  $t$ , and 0.05 contour line for the probability function.

case, although (1) is not normal in  $s$ , a rough numerical integration extended over the exterior of the contour line gives 0.052. Whether one would say that the samples drawn were significantly anomalous in their values of  $s$  and  $m$  if their representative points fell outside the contour line which contained 95% of the samples when judged by the values of  $s$  and  $m$ , is a matter of definition for the individual or of convention for statisticians as a group.

In figure 2 the diagram (for samples of 5) is transferred to the  $T$ - $t$ -plane. In the legend the probabilities that a point (sample) lie in each of the 10 regions are given. Outside the 0.05 contour line there is clearly considerable more chance of samples than 0.05, as is not unreasonable in view of the very



wide divergence of the frequency distribution from normal; a rough numerical quadrature indicates that 11% of the points lie outside the contour line.



Regions of the  $T$ - $t$ -plane connected with significance of mean, standard deviation and the ratio  $t$ , and 0.05 contour line for the probability function. The probability that  $(T, t)$  shall be in the various regions is

Region	I,	II,	III,	IV,	V,	VI,	VII,	VIII,	IX,	X
Probability	0.238,	0.8830,	0.0137,	0.0194,	0.0100,	0.0012,	0.0000,	0.0282,	0.0193,	0.0012

For samples of 3, 5, 9 ( $n = 2, 4, 8$ , respectively) the four-fold tables which correspond to the four regions separated by the line  $T = 1.96$  and that for  $t$  equal to its value for  $P = 0.05$ , are, respectively,

	$t < 4.303$ $t > 4.303$		$t < 2.776$ $t > 2.776$		$t < 2.806$ $t > 2.806$	
$T < 1.96$	0.9130	0.0370	0.9206	0.0294	0.9281	0.0219
$T > 1.96$	0.0370	0.0130	0.0294	0.0206	0.0219	0.0281

For these cases the correlation coefficients  $r$  and  $r'$  for the cases where  $T$  and  $t$  are both taken positive and where they are taken either positive or negative, and the corresponding correlation ratios  $\eta$  and  $\eta'$  for the regression of  $T$  on  $t$  are

	$n = 2$	$n = 4$	$n = 8$
$r$	0	0.756	0.899
$r'$	0	0.886	0.959
$\eta$	0.764	0.860	0.923
$\eta'$	0.921	0.952	0.973

Fisher has emphasized that what a fiducial argument does is to reject, or fail to reject, at a stated level of probability a hypothesis to which it is appropriate.<sup>4</sup> What we have here done is to discuss in some detail for the  $t$ -test the regression of the test  $T$  which one would apply if  $\sigma$  were known upon the value of  $t$  which is computed from the data.

<sup>1</sup> As  $t$  is what is known from observation it is the regression of  $T$  on  $t$  which is of interest rather than that of  $t$  on  $T$  for which the relations are linear, viz.,

$$t_{\text{mode}} = \sqrt{\frac{n}{n+1}} T; \qquad t_{\text{mean}} = \sqrt{\frac{n}{2} \frac{(\Gamma n/2 - 1/2)}{\Gamma(n/2)}} T.$$

<sup>2</sup> The value  $\nu_{pq}$  is the same whether  $T$  and  $t$  are taken in absolute value or are allowed to keep their signs, provided  $p + q$  is even, but if  $p + q$  is odd the value of  $\nu_{pq}$  is zero in the latter case.

<sup>3</sup> Student's distribution for  $t$  has no mean or second moment when  $n = 1$  and no second moment when  $n = 2$ ; note that (5) is imaginary for  $n = 1$  and zero for  $n = 2$ .

<sup>4</sup> R. A. Fisher has pointed out very precisely the rather specific conditions under which a fiducial argument may properly be given. See his notes: *Ann. Eugenics*, 4, 391-398 (1935) and *Ann. Math. Statis.*, 10, 383-388 (1939).



# PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

Volume 28

August 15, 1942

Number 8

Copyright 1942 by the National Academy of Sciences

---

## *DIFFUSION OF GENE PRODUCTS*

BY KARL SAX

ARNOLD ARBORETUM, HARVARD UNIVERSITY

Communicated June 24, 1942

The phenotypic expression of a gene may depend upon its relation to neighboring genes in the chromosome, the transmission of gene products from the nucleus to the cytoplasm and the diffusion of gene products from cell to cell. Although examples of position effect are rather limited, it is clear that the effect of a gene may be conditioned by neighboring genes. This action might be attributed either to the effect of adjacent genes in the production of gene products, or to the early interaction of products of adjacent or neighboring genes. There is abundant evidence that gene products may pass through the nuclear membrane.<sup>1</sup> Even during the "resting stage" of a cell the genes are active in controlling cellular development, as shown by the variation in growth of microspores in triploid *Tradescantias*.<sup>2</sup> The diffusion of gene products from cell to cell or even to relatively remote tissues is shown by the transplant experiments of Beadle, Ephrussi and Caspari in *Drosophila* and *Ephestia*.<sup>1</sup> Larval eye discs transplanted to abdomens of other larvae may be autonomous in pigment development or may be controlled by the host, depending on the genetic constitution of the implant and of the host. A remarkable case of a similar nature has been found in *Habrobracon* by Mrs. Whiting.<sup>3</sup> Eye mosaics may have a sharp line of cleavage between the colored segments, or the colors may merge and interact at the boundary of the mosaic areas, depending on the genetic composition of the mosaic males. It is evident that certain gene products are diffusible while others are not.

The development of the microspores in certain plants provides further evidence on the diffusion of gene products. In nearly all basic diploid plants the microspores rarely develop into pollen grains if a chromosome or even part of a chromosome is missing. Due to "non-disjunction" of bivalent chromosomes at the first meiotic division the daughter nuclei are either deficient or carry an extra chromosome. The deficient nuclei undergo the second meiotic division, because the cells have been condi-

tioned for one more division cycle, but the resulting deficient microspores fail to develop. The rare occurrence of  $N-1$  microspores which reach the stage of nuclear division has been attributed to non-disjunction at the second meiotic division.<sup>4</sup> In *Uvularia* the deficient microspores die if they become separated from their complementary hyperploid sister cells, but heat treatment during meiosis often causes the microspores from a pollen mother cell to remain attached. Barber<sup>5</sup> found that the deficient microspores in such attached cells undergo nuclear division and that the nuclear divisions in both the hypoploid and hyperploid microspores are synchronized. Apparently substances essential for growth and development are able to pass from hyperploid to deficient microspores only when the cells are attached. Barber has described similar cases in certain orchids.<sup>6</sup>

In the normal development of *Tradescantia* microspores there are conspicuous granules in the cytoplasm during early development. These granules gradually disintegrate and begin to disappear before the initiation of prophase of the nuclear division. In sterile microspores the granules persist and are present when nuclear division occurs in the neighboring normal cells. The sterile microspores fail to develop and are much smaller than normal microspores. According to Schmitt and Johnson<sup>7</sup> the granules in *Tradescantia* microspores are composed primarily of protein.

Occasionally *Tradescantia* microspores are found in which the chromosomes are separated by complete or partial cell wall formation. These dumb-bell shaped microspores have three chromosomes in each lobe and the lobes are separated by cell walls or by deeply constricted regions (figure 1). The aberrant microspores are about as large as normal cells, but the protein granules persist until metaphase. The granules had disintegrated more than those found in adjacent sterile microspores, but were very conspicuous when compared with the condition in normal microspores (figure 1). The origin of these aberrant microspores is obscure. They may be produced by non-disjunction at the first meiotic division, followed by incomplete separation of the two daughter cells of the deficient dyad, or they may be produced by the failure of the chromosomes to unite in a single nucleus at the telophase of a normal second meiotic division.

The development of attached half-microspores with only three chromosomes in each cell is not unexpected since deficient cells are conditioned for the following nuclear division. The persistence of the protein granules does suggest that the action of all chromosomes must occur to effect disintegration of the granules. Apparently gene products from the isolated groups of chromosomes cannot diffuse fast enough, especially when separated by a cell wall, to effect the complete disintegration of the granules. The growth of the cell is nearly or quite normal, however, as one would expect in view of Barber's results with attached microspores. Evidently the growth-promoting substances can pass through the thin walls of

attached microspores, but the products responsible for granule disintegration are inhibited by cell walls or perhaps even by cytoplasmic isolation. These results with plant microspores are in complete accord with the variation in diffusion of gene products found in *Drosophila* and *Habrobrachon*.

One of the most interesting cases of diffusion of gene products has been described in *Neurospora* by Dodge.<sup>8</sup> Two races when grown together produce a heterocaryotic mycelium containing nuclei from each parental

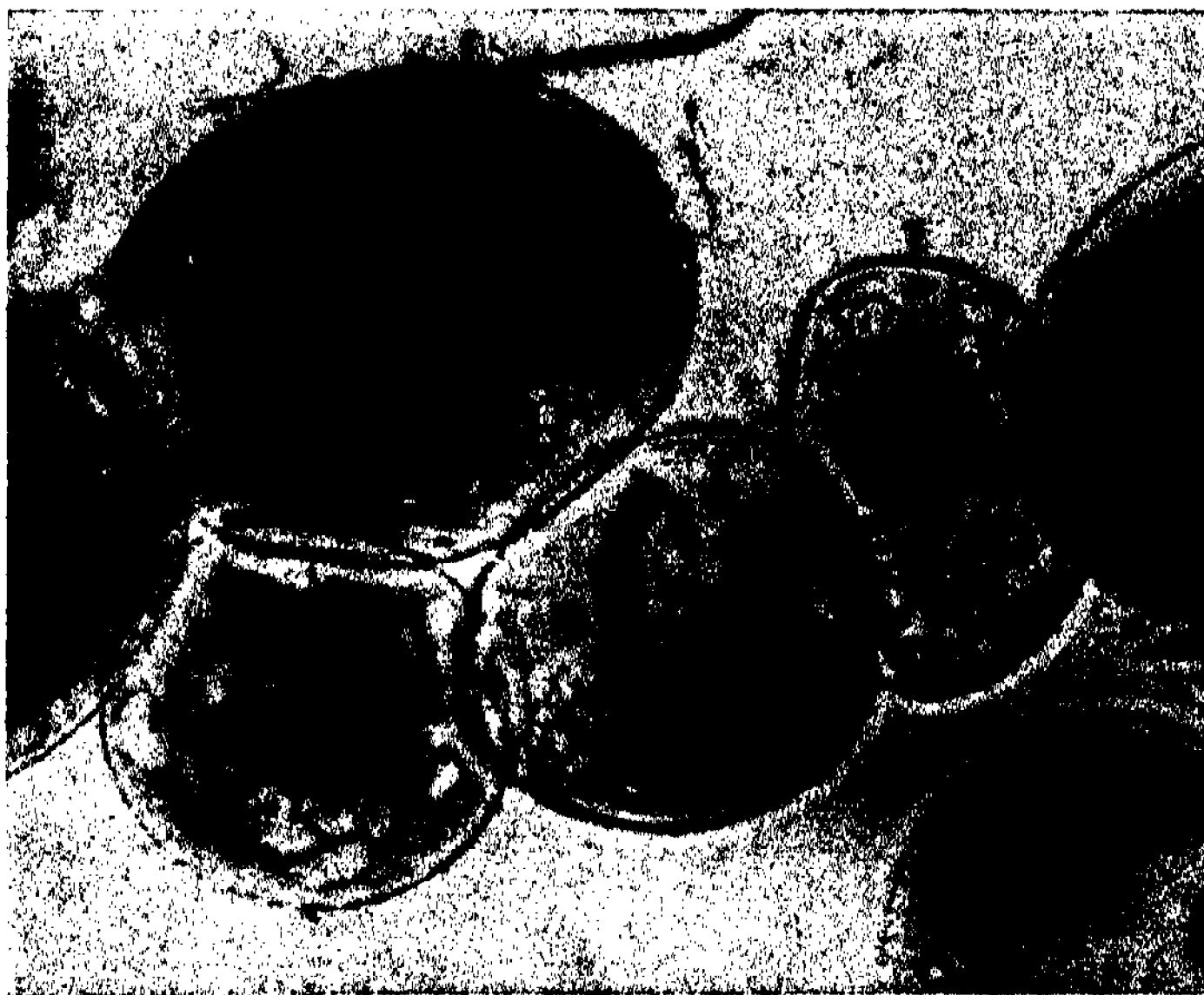


FIGURE 1

An aberrant *Tradescantia* microspore is shown with three chromosomes in each part of the dumb-bell shaped structure. Total growth is nearly normal, but the protein granules have persisted, due presumably to the isolation of gene products in the isolated chromosomes of the basic genom. A normal microspore is shown above and a sterile microspore to the right of the aberrant cell.

race. The resulting mycelium shows a great increase in growth although no nuclear fusion occurs. Dodge suggests that the growth substances produced by the nuclei of the two races supplement each other to produce the increased growth—a situation comparable to the hypothesis suggested by Robbins to account for heterosis in tomatoes. The results with *Neurospora* show that gene products necessary for growth need not be produced in the same nucleus in order to be effective, but can diffuse into the cytoplasm from different nuclei and unite in promoting increased vigor.

*Summary.*—Gene products essential for growth are able to diffuse between attached cells, but not between isolated microspores. The gene products necessary for disintegration of protein granules are unable to diffuse freely through a cell wall in aberrant *Tradescantia* microspores. This variation in the diffusion of gene products is in accord with the results found in insects.

<sup>1</sup> Wright, S., *Phys. Rev.*, 21, 487-527 (1941).

<sup>2</sup> Sax, K., *Genetics*, 22, 523-533 (1938).

<sup>3</sup> Whiting, A. R., *Coll. Net.*, 8, 29 (1933).

<sup>4</sup> Conger, A. D., *Jour. Heredity*, 31, 339-341 (1940).

<sup>5</sup> Barber, H. N., *Jour. Gen.*, 42, 245-257 (1941).

<sup>6</sup> Barber, H. N., *Ibid.*, 43, 97-103 (1942).

<sup>7</sup> Schmitt, F. O., and Johnson, G. T., *Ann. Mo. Bot. Gard.*, 25, 455-486 (1938).

<sup>8</sup> Dodge, B. O., *Bull. Torrey Bot. Club.*, 69, 75-91 (1942).

## THE EFFECT OF X-RAY STIMULATION ON THE BIOELECTRIC POTENTIALS OF THE AVIAN EGG\*

BY ALEXIS L. ROMANOFF AND ARTHUR A. BLESS

AGRICULTURAL EXPERIMENT STATION, CORNELL UNIVERSITY, AND DEPARTMENT OF PHYSICS, UNIVERSITY OF FLORIDA

Communicated June 12, 1942

*Introduction.*—The electrical activity of living systems has recently received a great deal of attention from a number of investigators. It is recognized that bioelectrical potentials are not merely an accidental by-product of the activity of the living organisms, but are probably conditioned by the organism and, according to some observers, profoundly influence its living processes. The electrical potentials and the electrical fields are created by ions and their spacings which are "the irreducible relatedness of the components of living things," as Northrop and Burr<sup>1</sup> expressed it.

The vital activity of the blastoderm of the hen's egg has been investigated recently by Romanoff and Cottrell.<sup>2</sup> The results have shown that the potential difference between the blastoderm and the albumen of fresh fertile eggs is much larger than the potential of infertile eggs. The electrical activity of the blastoderm is therefore an indication of the vital activity of the organism. In view of the importance of this criterion of vital activity it seemed interesting to determine the physical factors that affect the potential, and the bearing these factors have on the later life of the embryo. This paper deals in particular with the effects of x-rays on the bioelectric potential.

*Experimental Methods.*—About 1200 fresh White Leghorn eggs<sup>3</sup> were

subjected to x-ray doses ranging from 8 to 5000 r units. The radiation was supplied by a Coolidge, water-cooled, tungsten target tube with a  $45^\circ$  focus. The tube was operated at 55,000 volts and 10 milliamperes. The radiation was filtered with 1.2 mm. of aluminum. Under these conditions the band of radiation was fairly narrow, extending from 0.24 to 0.36 Angström units.<sup>4</sup> Output of the tube was approximately 600 r units per minute at a distance of 14 cm. from the target.

The apparatus used for the measurements of the bioelectric potentials has been described elsewhere.<sup>2</sup> The sensitivity of the system was about 20

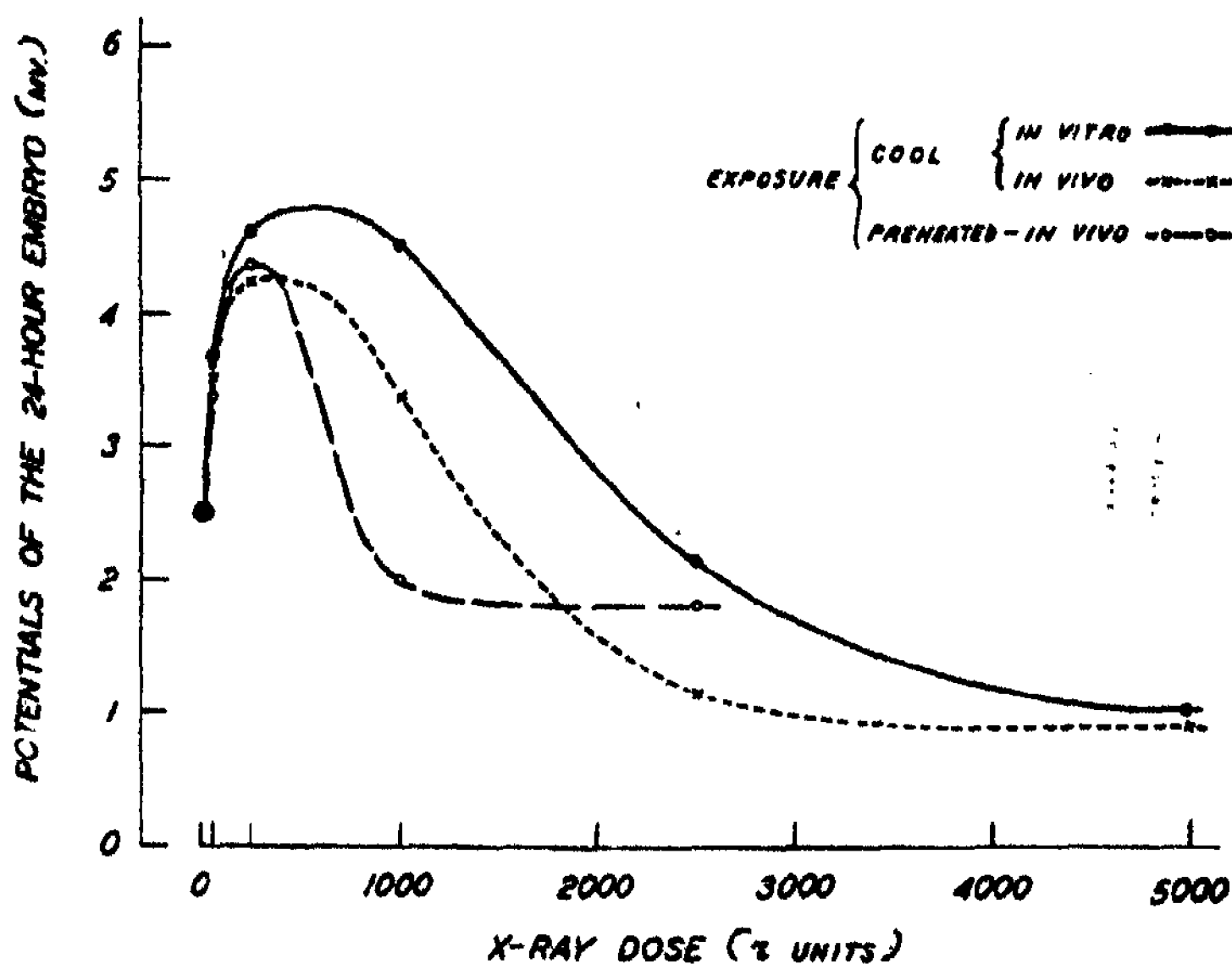


FIGURE 1

The effects of various doses of x-rays on the bioelectric potential of the 24-hour chick blastoderm.

microvolts per millimeter of scale deflection on a wall galvanometer. This sensitivity proved to be ample for the purpose.

The eggs were irradiated *in vitro*, with the yolks submerged in the albumen and held in small beakers by glass rings, and also *in vivo*. One half of the eggs irradiated *in vivo* were exposed to x-rays when cold, while the other half were preheated for three hours before exposure. The potential differences between the blastoderm and the albumen and also between the yolk and the albumen of the developing eggs were measured for 24 hours at 6-hour intervals.

**Results.**—Figure 1 shows the variation of the potential of the 24-hour



embryos as a function of the x-ray dosage. It is evident from this figure that x-rays influence profoundly the vital activity of the embryo. Small doses exert a stimulating effect with the maximum stimulation occurring at about 250 r units. Larger doses inhibit vital activity. This agrees with the experiments on the early growth of the blastoderm as a function of x-ray exposure.<sup>5</sup> At very large doses the x-ray action becomes lethal with the bioelectric potential reduced greatly, approaching the value of non-fertile eggs. The behavior of the bioelectric potential under the influence of x-rays is thus similar to the behavior of plants which are stimulated by small doses and are inhibited by large exposures to x-rays.<sup>6, 7, 8</sup>

It is also evident from figure 1 that the bioelectric potential of eggs irradiated to inhibitive doses of x-rays when preheated was lower than the potential of those irradiated when cold. This indicates that the biological effect of x-rays on an organism is greatest when the activity of the organism is high, which again is in agreement with similar experiments on plants.<sup>9, 10</sup> There is, however, an important difference. The effect of x-rays on sprouting seeds is 15–20 times the effect of x-rays on the same seeds irradiated in the dormant state.<sup>11</sup> On the other hand the effect of x-rays on the preheated eggs is only a few per cent greater than the effect of x-rays on cold eggs. The difference is doubtless due to the fact that the blastoderm of fresh eggs is probably not in a state of complete dormancy, as is the dry seed.

The range of doses exerting stimulating effects is greater for cold eggs, both *in vitro* and *in vivo*, than for preheated eggs. This is to be expected because the effects of x-rays on the inactive embryo are much smaller than the effect on the active embryo; naturally the range of doses exerting stimulating action on cold eggs is larger.

The effect of x-rays on bioelectric potentials of preheated fresh and incubated eggs is shown in figure 2. From this figure it is evident that the range of stimulating doses of x-rays is much wider in fresh eggs than in those incubated for 24 hours (Fig. 2 (A)). The relative changes in potential (Fig. 2 (B)) were observed to be greatest in the fresh eggs and least in those incubated six hours. A noticeable increase in potential of eggs incubated 24 hours would indicate that the general detrimental effect of x-rays becomes apparent only after a certain time<sup>5, 12</sup> and is followed by a period of physiological recovery. The lag of the effect may be explained in the following way.

The biological effects of x-rays are no doubt a result of the physical absorption of x-rays and the accompanying photoelectric and ionization action. The full effect of the photoelectric action with the consequent chemical changes and the biological effects become apparent only after a more or less extended period. On the other hand, while the biological effects occasioned by the x-rays are taking place, the inertial and defensive

forces of the organism combat these effects of x-rays and help the organism to recuperate from these effects. It is evident, therefore, that at a certain time after exposure the effects will be a maximum, after which the defensive and the recuperative forces will diminish the effectiveness of the x-ray doses, unless the action is lethal.

The potential of the blastoderm seems to be greatest after about 18 hours of incubation. This, however, does not mean that the activity of the

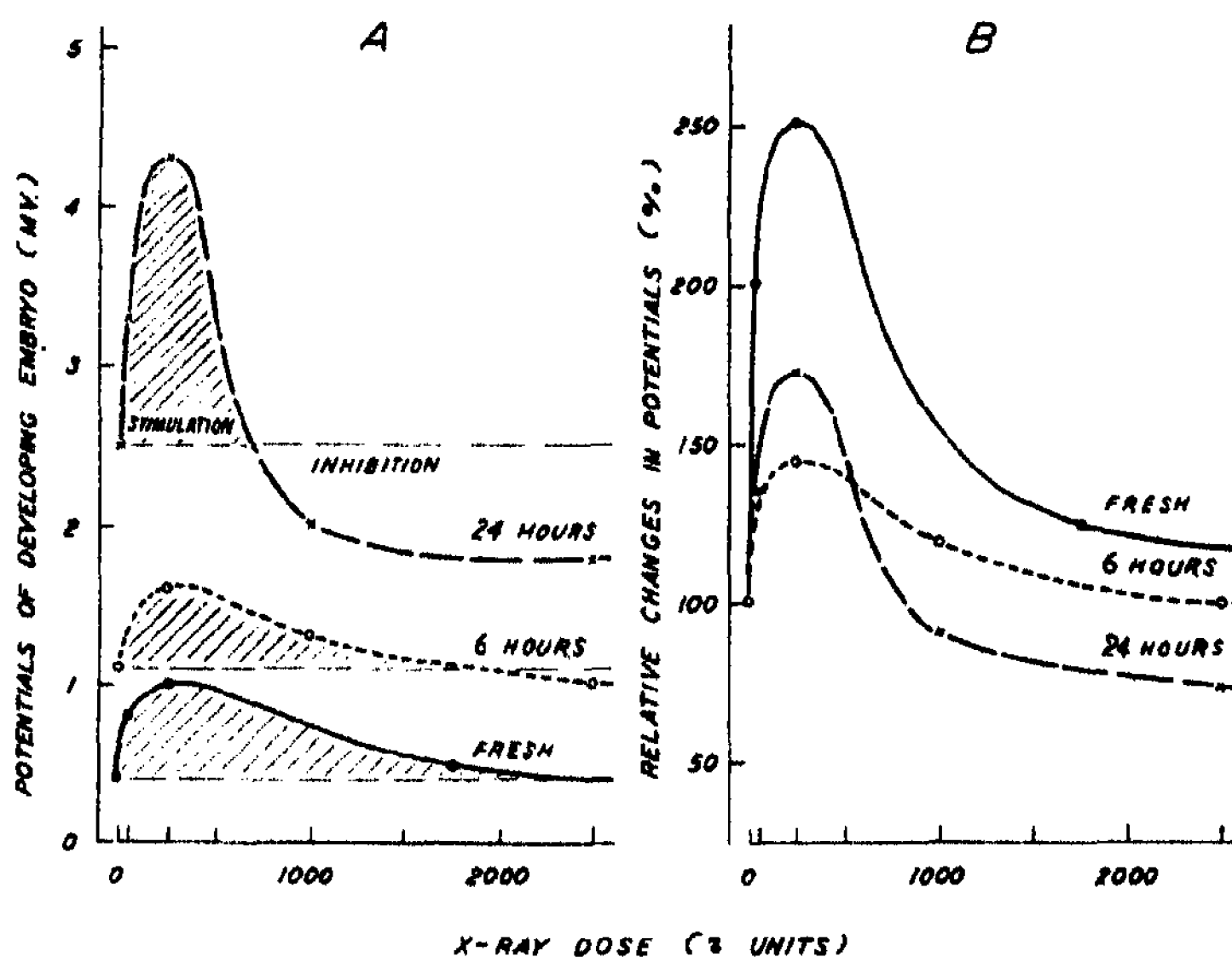


FIGURE 2

The effect of exposure of preheated eggs to various doses of x-rays on the bioelectric potential of the chick blastoderm in its early stages. *A*, potential differences in millivolts of the blastoderm of fresh eggs, and incubated six and twenty-four hours. *B*, relative changes in potential differences (control considered 100 per cent). The eggs were preheated for 3 hours at incubating temperature.

organism diminishes beyond this period. The growth of the blastoderm proceeds at a very rapid rate,<sup>5</sup> but because of the physiological differentiation it is not possible to speak of the potential of the blastoderm as a whole. Each part of the blastoderm definitely gives a different electrical potential. This has been well demonstrated by Burr and Hovland<sup>13</sup> in the study of cephalo-caudal potential of the chick embryo.

The potential difference between the yolk and the albumen is usually very small. This, however, is not always the case. A small region, about one square centimeter, was found on a number of yolks, which was charac-

terized by an extraordinarily high potential, higher than that of the early blastoderm, as is shown in the table below on a group of unexposed eggs.

	BLASTODERM	MAIN BODY OF YOLK	SMALL REGION OF YOLK
Fresh fertile eggs	0.5 mv.	0.8 mv.	2.7 mv.
Incubated 6 hours	1.3 "	0.7 "	2.9 "
Incubated 24 hours	3.0 "	0.5 "	1.3 "
Infertile	0.3 "	0.2 "	1.5 "

While the frequency of such cases was not very great, the presence of such regions in the yolk makes it undesirable to measure the potential of the blastoderm in intact eggs between two points on the opposite poles of the equatorial plane as has been done by Vorontsov and Serguijevski.<sup>14</sup> The potential attributed to the blastoderm may actually be due to the yolk. In most cases the high potential area was slightly darker than the rest of the yolk and the line of separation of the two regions served as a boundary of the high potential area. The potential changed abruptly when this boundary was crossed. In some cases the high potential region showed no visible differences in coloring.

Since in the vast majority of cases the potential difference of the yolk remains unchanged under various exposures to x-rays, it is safe to conclude that the electrical effects of x-rays are confined to the living portion of the egg—the blastoderm.

In some very few cases the potential between the blastoderm and the albumen or between the yolk and the albumen was reversed. The origin of this reversal is not at all clear.

*Summary.*—These experiments indicate that:

1. X-ray doses of about 250 r units applied to fresh eggs stimulate vital activity of the developing embryo with the increase in the potential difference between the blastoderm and the albumen.
2. The range of stimulating doses is much greater for cold eggs, than for preheated eggs.
3. Large doses of x-rays inhibit embryonic development and diminish the potential difference.

\* This work was done at the Department of Physics, University of Florida.

<sup>1</sup> Northrop, F. S. C., and Burr, H. S., *Growth*, 1, 78-88 (1937).

<sup>2</sup> Romanoff, A. L., and Cottrell, C. L., *Science*, 90, 471-472 (1939).

<sup>3</sup> The eggs were supplied by the Agricultural Experiment Station Farm, University of Florida.

<sup>4</sup> Bless, A. A., *Proc. Nat. Acad. Sci.*, 23, 194-196 (1937).

<sup>5</sup> Bless, A. A., and Romanoff, A. L. (in press).

<sup>6</sup> Koernicke, M., *Jahrb. Wiss. Bot.*, 56, 416-430 (1915).

<sup>7</sup> Shull, C. A., and Mitchell, J. W., *Plant Physiol.*, 8, 287-296 (1933).

<sup>8</sup> Bless, A. A., *Ibid.*, 13, 209-211 (1938).

- <sup>9</sup> Komuro, H., *Zeitsch. Krebsforsch.*, 22, 199-209 (1925).  
<sup>10</sup> Wernhart, K., *Compt. Rend. Soc. Biol.*, 96, 1090-1091 (1927).  
<sup>11</sup> Stadler, L. J., *Jour. Hered.*, 21, 3-19 (1930).  
<sup>12</sup> Marshak, A., *Proc. Nat. Acad. Sci.*, 23, 362-369 (1937).  
<sup>13</sup> Burr, H. S., and Hovland, C. I., *Yale Jour. Biol. and Med.*, 9, 247-258 (1937).  
<sup>14</sup> Vorontsov, B. S., and Serguijevski, M. V., *Probleme der Tierzucht*, No. 6, Moscow (1933).

## NUCLEIC ACID STORAGE IN THE TOAD'S EGG

BY THEOPHILUS S. PAINTER AND A. N. TAYLOR

UNIVERSITY OF TEXAS

Communicated June 29, 1942

From chemical analyses it has been known for some time that the unsegmented eggs of a number of different animals carry a high proportion of the nucleic acid which is present at the time larvae hatch (e.g., frog 28%, sea urchin 100%).<sup>1</sup> Brachet<sup>2</sup> has shown, in the case of the sea urchin, that this nucleic acid is initially in the ribose form (ribonucleic acid) which is partially transformed into the desoxy ribose type (thymonucleic acid or chromatin) as development goes forward. The writers have been studying the cytological mechanisms by which this nucleic acid is laid down in the egg cytoplasm.<sup>3, 4</sup> The present report deals with the oöcytes of the toad (*Bufo valliceps*), a form of especial interest because of the enormous germinal vesicle and the presence of the so-called "lampbrush" chromosomes. In the past, both of these features have been interpreted in various ways as contributing to the nucleic acid reserve of the egg cytoplasm.

**Methods.**—Fresh ovaries were preserved either in Nawaschin's fluid and the sections stained by Feulgen's Nucleal method, or fixed in Helly's fluid and stained by Unna's methyl green-pyronin mixture, both before and after treatment with a ribonuclease enzyme.

**Observations.**—Figure 1 is a camera lucida drawing showing in the wall of the ovary an oögonial cell in prophase and two very young oöcytes in the bouquet stage (one nucleus drawn to show the chromosomes, the other the surface of the nuclear wall). In the oögonium all of the chromatin is confined to the prophase chromosomes. In the oöcytes, on the other hand, the chromatin is found in two types of structures: (a) the chromosomes which have the loop form characteristic of the synizesis stage, and (b) free chromatin granules which lie against the inside of the nuclear wall and are most numerous on the side of the nucleus opposite that towards which the ends of the chromosome loops are oriented. This distribution of the nuclear contents makes it possible to observe that there is no connection at this

time between the chromosomes and the free granules, an important point. The subsequent history of the oöcyte is best followed by considering the behavior of the chromosomes and the free chromatin granules separately.

At the bouquet stage the chromosomes appear to be made up of granules embedded in unstained material, the matrix, which imparts a rather fuzzy outline to the whole (Fig. 1). As the egg nucleus, or germinal vesicle, grows in size the chromosomes increase in length by uncoiling and reach their maximum extension when the nucleus is about 40 or 50  $\mu$  in diameter. With a further increase in nuclear size the chromosomes occupy only a limited area within the germinal vesicle. Turning to details, when the nucleus is about double its initial diameter (10  $\mu$ ) at the bouquet stage, the individual chromosomes (Fig. 2 (a)) are still relatively thick and granular and quite jagged in outline. Part of the latter is due to the zig-zag arrangement of the chromomeres (chromioles) and part to lateral extensions of the achromatic matrix into the nuclear sap, as we interpret it forming the side branches which characterize the lampbrush chromosomes. The chromosomes are not uniform in staining reaction. There are localized areas of different chromosomes which are thicker, because the chromatic granules are larger, as indicated in figure 2 (a). These heterochromatic areas persist throughout the growth period. In nuclei about 30  $\mu$  in diameter (Fig. 3 (a)) the chromosomes are thinner and longer and the chromomeres appear to be arranged in a zig-zag manner in a mid-focal plane due to the uncoiling of the chromomeric threads of which each chromosome is composed. The side branches, which do not color with Feulgen's stain, are more pronounced. With further extension in length the Feulgen positive chromomeres assume a linear order, a condition which is retained until a contraction sets in prior to the breakdown of the germinal vesicle wall (Fig. 4). In the later stages (from 30  $\mu$  on) it is obvious that each chromosome is made up of two threads held together by chiasmata, but we have been unable to observe any split in the two threads, which the presence of chiasmata would imply.

After Feulgen's stain the chromosomes in later stages are rather difficult to observe but the chromomeres do not lose their purple color, contrary to the report of Koltzoff<sup>8</sup> and others. In the toad the side branches of the chromosomes are never conspicuous and are best seen with oblique illumination. When sections are heavily overstained with haematoxylin, however, we get the classic lampbrush image except for the absence of loops which we have not seen in toad eggs fixed in Nawaschin's or Helly's fluids.

So far we have been unable to distinguish oöcytes earlier than the bouquet stage and have not observed just how or when the chromatic granules are separated from the meiotic chromosomes. The former lie just within the nuclear wall, are somewhat variable in size (best shown in tangential section of the lower nucleus in figure 1) and are most numerous on the side

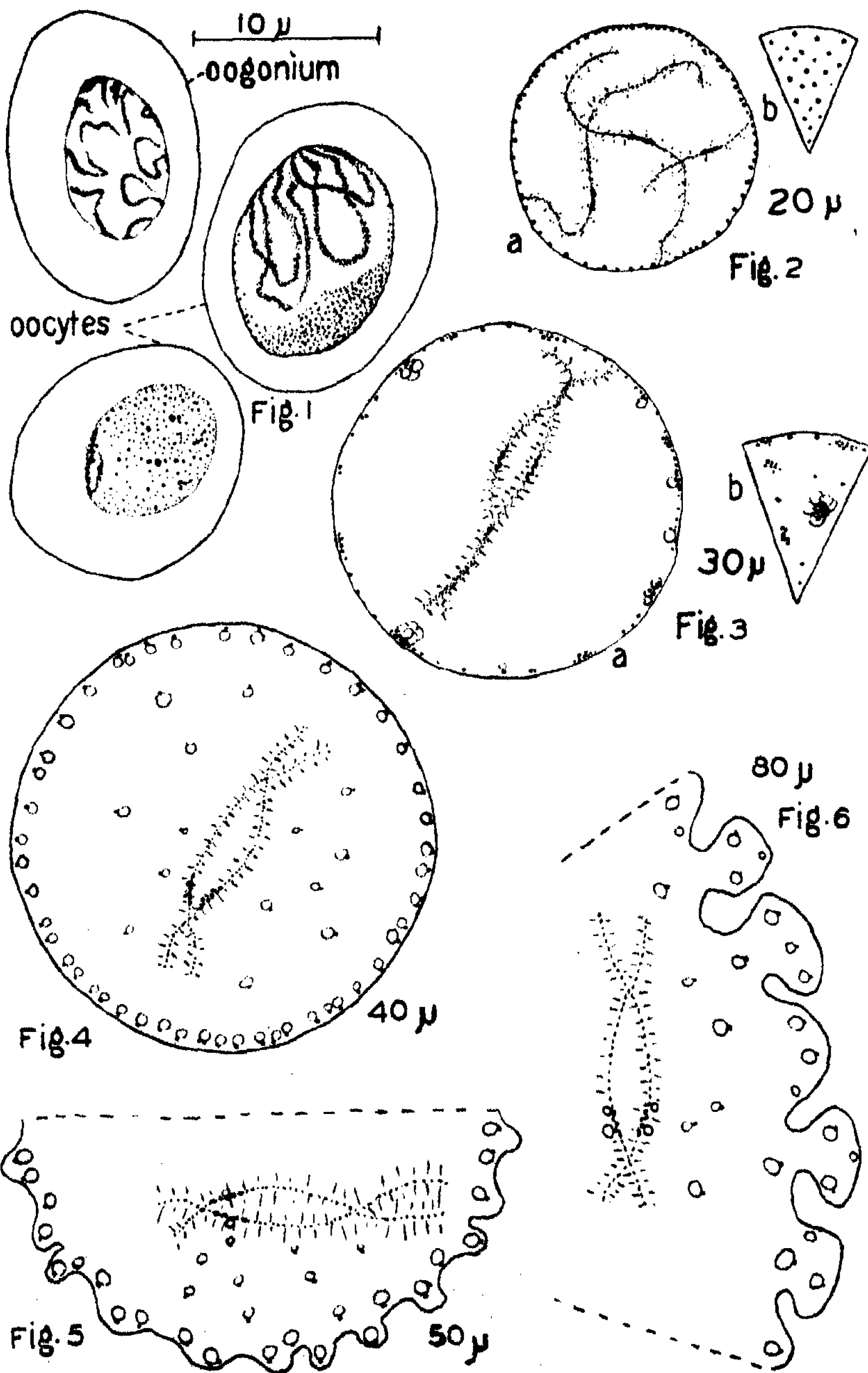


Figure 1 is a camera lucida drawing and the remaining figures are semi-diagrammatic.

of the nucleus opposite to which the free ends of the chromosome loops are oriented. As the germinal vesicle grows in size there is little change in the granules until a nuclear diameter of about  $30\ \mu$  is reached when they show a pronounced tendency to associate into rather large clumps (Figs. 3 (a) and 3 (b)). This behavior suggests that the granules are heterochromatin because several workers have noted that heterochromatin from different chromosomes tends to synapse at about this point in meiosis. About this time small nucleoli begin to arise in association with the granules; clusters about the clumps, or single nucleoli against single granules. A little later the granular clumps break up and spread out and thus the whole inner surface of the nucleus becomes lined with small nucleoli with the chromatin organizers directed outward (Fig. 4). Usually each nucleolus shows only one chromatin granule attached to it but occasionally there may be two or more. The nucleoli proper are fairly uniform in diameter at first; there are several hundreds of them and they do not take Feulgen's stain, though with haematoxylin they stain intensely.

When the germinal vesicle reaches a diameter of about  $60\ \mu$  its wall, which previously has been smooth in outline becomes wavy and soon shows numerous finger-like pseudopodia or processes each of which contains one or more nucleoli (Fig. 5). We now note a good deal of variation in the size of nucleoli and the chromatin organizers begin to disappear as the nucleoli diminish in size, as if they are being absorbed (Fig. 6). In the meantime new nucleoli are seen to be arising in connection with the heterochromatic areas of the chromosomes. After they become detached, each shows the characteristic Feulgen positive granule directed towards the nuclear wall. These nucleoli appear to migrate to the surface of the nucleus. In older germinal vesicles the new crops of nucleoli tend to lie in concentric areas about the chromosomes which suggests that their production may be cyclical. We have seen no evidence for a bodily extrusion of nucleoli into the egg cytoplasm.

*Experiment with Ribonuclease.*—The observations recorded above were, for the most part, reported by us in December, 1940. In the same year but unknown to us (because the journal did not reach us until in 1941), Brachet<sup>6</sup> reported that he had used Feulgen's technique on the eggs of frogs and found the Feulgen positive granules associated with the nucleoli. In addition, using a heat stable ribonuclease, which he isolated from pancreas, Brachet was able to prove that both the nucleoli and the cytoplasm of the frog's egg are very rich in ribonucleic acid. We have repeated these experiments using a crystalline ribonuclease kindly furnished us by Dr. Kunitz, and can confirm Brachet's findings. When toad eggs are preserved in Helly's fluid and the sections stained in Unna's methyl green-pyromin mixture both the nucleoli and the cytoplasm stain a brilliant red. Some red-staining material is present in very young oöcytes and in older eggs there is a very large



amount of it with a rather sharp gradient, the color being most intense next to the nuclear membrane and shading off towards the cortex. If, now, sections are first treated with ribonuclease for a few hours, and then stained with Unna's mixture, the red color no longer shows in either the nucleoli or the cytoplasm. In control slides treated with the inactivated enzyme the staining is the same as in untreated slides. Except for a loss of staining capacity, nucleoli are not visibly affected by ribonuclease. Brachet interprets these results to mean that the red color is due to the presence of ribonucleic acid which is removed by the ribonuclease treatment. Obviously then, the heterochromatic granules and the nucleoli associated with them are part of a mechanism for the laying down of ribonucleic acid in the cytoplasm of the toad's or frog's egg.

*Discussion.*—Before Feulgen's technique was developed for the identification of chromatin (thymonucleic acid) it had been commonly reported that in eggs with large germinal vesicles the chromosomes, the numerous nucleoli and various unidentified nuclear inclusions, stain deeply with basic dyes. And a number of observers, noting that the amount of basophilic material which finally entered the first polar spindle (as chromosomes) was only a small fraction of the total seen earlier in meiosis, suggested that with the breakdown of the germinal vesicle wall this excess nuclear material would be available to the developing embryo (Conklin,<sup>6</sup> Godlewski,<sup>7</sup> Koltzoff<sup>8</sup>). The present study, however, indicates that there is a specific cytological mechanism in the toad (and frog) which begins to deposit ribonucleic acid in the cytoplasm of the oöcyte soon after it is differentiated and continues to function through the months required to build up the mature egg. While large amounts of nuclear material are set free in the cytoplasm when the germinal vesicle breaks down, this appears to contain very little nucleic acid of either the ribose or desoxy ribose type. This conclusion is supported by the observations of Caspersson and Schultz<sup>9</sup> that in the sea urchin the mature germinal vesicle is relatively poor in pyrimidine bases while the cytoplasm just outside the nuclear wall is very rich in them.

Within the past five years a good deal of evidence has come to light which indicates that there is an intimate and causal relationship between heterochromatin, nucleolar formation and the synthesis of ribonucleic acid in the cytoplasm of the cell.<sup>9, 10</sup> Broadly speaking, wherever cells undergo rapid protein synthesis, as in growth or secretory activity, there is usually an abundance of ribonucleic acid in the cytoplasm and prominent nucleoli in highly basophilic nuclei. In the toad this cytological and dynamic relationship stands out very clearly because the heterochromatin, which is the main source of the basophily in growing or secretory cells, is not inextricably mixed with the chromosomes but is segregated for the most part into discrete granules entirely removed from the chromosomes, and instead of one



or several nucleoli there are hundreds of them. Here each nucleolar complex consists of one or more heterochromatin granules, the main body of the nucleolus probably protein in nature,<sup>9</sup> and some ribonucleic acid the removal of which with ribonuclease does not morphologically affect the body of the nucleolus. It is clear that ribonucleic acid can be synthesized, or converted (from desoxy ribose) deep within the nucleus, and with the migration of the nucleoli to the nuclear wall and their absorption this is somehow transferred to the cytoplasm. But the localization of the heterochromatic granules initially on the inner wall of the nucleus, the invariable orientation of the granules towards the cytoplasm and the sequence of changes which occur at this site all suggest that most of the synthesis or conversion is occurring at the interface of the nucleus and cytoplasm where an abundance of surface energy is available.

Our observations show that the lampbrush chromosomes of the toad are normal and typical meiotic chromosomes in structure and behavior. This fact is stressed because earlier one of us (Painter<sup>8</sup>) postulated that lampbrush chromosomes were the result of some sort of reduplication process. Ordinarily, a great increase in nuclear volume is accompanied by the growth and division of the contained chromosomes, forming the giant chromosomes in the salivary glands of Diptera or highly polyploid nuclei as in the larval cells of many insects which grow by endomitosis. Koltzoff<sup>8</sup> thought that the side branches of lampbrush chromosomes represent reduplications of the primary "genonema" which are thrown off before polar spindle formation and which merge with the cytoplasm when the germinal vesicle breaks down. But since the granules which the side branches appear to enclose contain neither thymonucleic acid, as shown by Feulgen's stain, nor ribonucleic acid, as shown by Unna's stain, there is little to support this concept. On the other hand, with the increase in nuclear volume in the toad's egg there is a great increase in the amount of heterochromatin and in the number of nucleoli which form in association with the heterochromatic granules. If these nucleolar organizers are genetically the same as those which form nucleoli in ordinary somatic cells, then we may say that the germinal vesicle of the toad is highly polyploid in nucleolar organizers but otherwise lampbrush chromosomes are normal meiotic structures.

Perhaps the most important aspect of the present study is the striking demonstration that chromatin can exist and function within the nucleus apart from the chromosomes. And thus we are able to add another characteristic to the already long list of attributes of heterochromatin recently summarized by Darlington,<sup>11</sup> and the rôles this plays in the dynamic activities of cells.

<sup>1</sup> Needham, J., *Chemical Embryology*, II (1931).

<sup>2</sup> Brachet, J., *Arch. Biol.*, **44**, 519-576 (1933).

- <sup>3</sup> Painter, T. S., *Proc. Nat. Acad. Sci.*, 26, 95-100 (1940).  
<sup>4</sup> Painter, T. S., and Taylor, A. N., *Anat. Rec. Suppl.*, 78, 84 (1940).  
<sup>5</sup> Brachet, J., *Arch. Biol.*, 51, 151-165 (1940).  
<sup>6</sup> Conklin, E. G., *Jour. Exp. Zool.*, 12, 1-88 (1912).  
<sup>7</sup> Godlewski, E., *Arch. f. Entw. Mech.*, 26, 278-328 (1908).  
<sup>8</sup> Koltzoff, N. K., *Biol. Zhurnal*, 7, 3-45 (1938).  
<sup>9</sup> Caspersson, T., and Schultz, J., *Proc. Nat. Acad. Sci.*, 26, 507-515 (1940).  
<sup>10</sup> Schultz, J., Caspersson, T., and Aquilonius, L., *Ibid.*, 26, 507-515 (1940).  
<sup>11</sup> Darlington, C. D., *Nature*, 149, 66-68 (1942).

## ON THE PERSEUS CLUSTER OF NEBULAE

BY F. ZWICKY

NORMAN BRIDGE LABORATORY OF PHYSICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated June 17, 1942

The Perseus cluster of nebulae (R.A.  $3^h 15^m$ , Dec.  $+41^\circ 15'$ , 1930; gal. long.  $118^\circ$ , lat.  $-13^\circ$ ) according to Hubble and Humason<sup>1</sup> contains about 500 nebulae scattered over an area nearly  $2^\circ$  in diameter and lies at a distance of 11 million parsecs. As is the case with most clusters investigated, counts of nebulae made on photographs which were obtained with the 18-inch Schmidt telescope show that the Perseus cluster is considerably larger than was originally derived from the photographs taken with the large reflectors whose very severely restricted fields make them unsuitable for the efficient analysis of objects subtending large angles. For quite a different reason the analysis of the spatial distribution of the nebulae in the Perseus cluster with the 18-inch Schmidt telescope also presents considerable difficulties. The cluster is projected on a field of the Milky Way so rich in stars, that, because of the small scale of the telescope, the blurred images of close pairs and groups of stars may easily be mistaken for extragalactic nebulae. This error can partly be avoided by taking a number of well-focused photographs while the telescope is being drifted slightly in a different direction for each photograph. Groups of stars are likely to betray themselves by an image containing sharp streaks such as should not be expected in a drifted image of a nebula. Also, the precaution was taken to identify all of the nebulae involved several times through a repeated analysis conducted during several years and making use of many photographs taken on various emulsions. It is therefore felt that the results presented here can be viewed with more confidence than might have been originally hoped for.

In figure 1 the distribution of nebulae brighter than about the photographic magnitude  $m_p = 16.5$  over a field of approximately  $9^\circ$  in diameter around the Perseus cluster is shown.

According to Hubble's investigations the Perseus cluster is seen through a large "window" of relatively high transparency in the galactic dust clouds. Lanes and pockets of heavy obscuration surround the cluster on the east, north and west sides. One dark pocket in which no nebulae and relatively few stars can be seen is marked *DP* in figure 1. The galactic plane lies in the direction NNE about thirteen degrees from the center of the

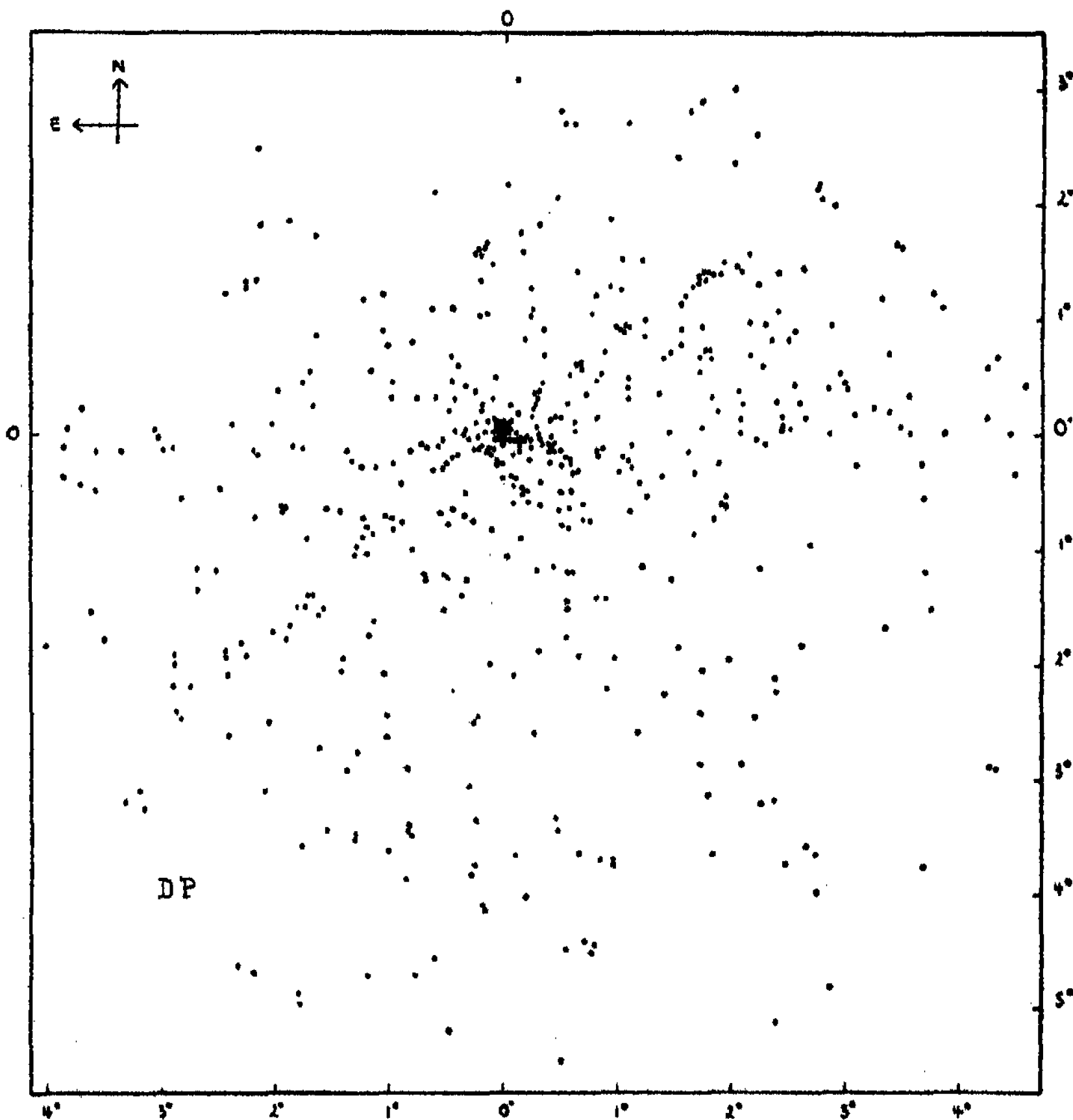


FIGURE 1  
The Perseus cluster of nebulae.

cluster. Since in this direction at distances larger than five degrees the obscuration seems to increase rapidly, we have counted the nebulae in complete rings around the center of the Perseus cluster only to a radius of 200' of arc. In table 1 are tabulated the total numbers  $n_r$  of nebulae contained in successive rings of 5' of arc and 10' of arc width, respectively. The last

TABLE 1  
COUNTS OF NEBULAE IN THE PERSEUS CLUSTER IN RINGS WHOSE RADII DIFFER BY 5'  
OF ARC AND 10' ARC, RESPECTIVELY

NO. OF RING	QUADRANTS				TOTAL $n_r$	$N_r$ /SQ. DEG.
	NE	NW	SW	SE		
1	4	3	2	3	12	557
2	5	3	4	2	14	217
3	2	2	5	7	16	149
4	4	2	4	5	15	99
5	1	3	5	3	12	62
6	2	4	6	1	13	55
7	4	2	7	2	15	54
8	1	3	4	4	12	37
9	2	0	4	3	9	25
10	1	4	2	3	10	24.4
11-12	1	6	7	4	18	19.0
13-14	5	6	8	2	21	18.7
15-16	4	3	4	7	18	13.8
17-18	3	8	3	8	22	15.0
19-20	5	8	4	5	22	13.4
21-22	4	5	3	3	15	8.3
23-24	2	11	6	3	22	11.0
25-26	2	9	6	6	23	10.7
27-28	2	7	2	8	19	8.2
29-30	2	10	2	5	19	7.6
31-32	4	9	2	4	19	7.1
33-34	1	5	3	3	12	4.22
35-36	1	6	3	6	16	5.30
37-38	1	7	1	8	17	5.33
39-40	1	2	4	4	11	3.27
<hr/>						
Total	64	128	101	109	402	
41-42	..	...	3	3	6	3.40
43-44	.	..	2	9	11	5.93
45-46	.	...	6	10	16	8.25
<hr/>						
			SOUTH			
47-48			4		4	3.95
49-50			5		5	4.73
51-52			2		2	1.82
53-54			0		0	0.00
55-56			5		5	4.22
57-58			4		4	3.25
59-60			4		4	3.15
<hr/>						
Total number of nebulae counted					459	

column gives the average numbers of nebulae  $N_r$  per square degree in these rings. The area inside of the first circle is  $1/46.4$  square degrees.

A total of 402 nebulae was counted within a radius of 200' of arc from the center of the cluster. Within the fluctuations to be expected these nebulae

distribute themselves uniformly over the four quadrants, except that perhaps the NE quadrant shows a slight deficiency in the number of nebulae, which is presumably due to the increasing effects of obscuration on approaching the galactic plane. From the general appearance of the stellar field it seems safe to count nebulae in the SW and the SE quadrants to a distance of  $230'$  of arc and in the south quadrant bounded by the SW and SE directions to a distance of  $300'$  of arc and beyond from the center of the cluster. At a distance of about  $250'$  of arc south of the center the cluster nebulae here considered merge with the field nebulae which appear at an average density of about 3.0 nebulae per square degree. According to the counts presented here the Perseus cluster therefore has an angular diameter of at least  $8^\circ$  of arc and an actual diameter of at least  $1.5 \times 10^6$  parsecs or five million light years.

According to Hubble<sup>2</sup> the average number  $N$  of nebulae per square degree in the general field is given by

$$\log_{10} N = 0.6m_L - 9.1. \quad (1)$$

In our case the limiting photographic magnitude  $m_L$  included in the counts is about  $m_L = 16.5$ . This according to equation (1) corresponds to  $N = 6.3$  nebulae per square degree. From our observations we obtained  $N' = 3.0$  nebulae per square degree instead, a result which according to (1) is equivalent to the average number over the unobscured parts of the sky of nebulae per square degree whose apparent photographic magnitude is smaller or equal to  $m_L' = 15.95$ . Since on Palomar Mountain the Perseus cluster can be observed within ten degrees of the zenith no appreciable zenith distance correction is to be introduced. If we therefore assume that the unobscured field of nebulae around the Perseus cluster is comparable to the average field of nebulae all over the unobscured parts of the sky we arrive at a value

$$\Delta m = m_L - m_L' = 0.55 \text{ magnitudes} \quad (2)$$

for the local obscuration. This value is presumably smaller than the actual value because it is likely that the Perseus cluster is imbedded in the very large cloud of nebulae extending from Andromeda over Pisces into Perseus, a fact which would result in a local value for  $N$  larger than the average all over the sky and in a value for  $\Delta m$  larger than given by (2). Of the 460 nebulae counted about 360 nebulae are physical members of the Perseus cluster. Since these nebulae are all brighter than about the absolute magnitude  $M = -14.3$  the total population of the Perseus cluster presumably includes well over one thousand nebulae.

<sup>1</sup> Hubble, E., and Humason, M. L., *Astrophys. Jour.*, **74**, 43 (1931).

<sup>2</sup> Hubble, E., *Ibid.*, **79**, 70 (1934).

## A REVERSIBLE GROWTH INHIBITION OF ISOLATED TOMATO ROOTS\*

BY JAMES BONNER

WILLIAM G. KERCKHOFF LABORATORIES, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated June 17, 1942

*Introduction.*—It is known that certain bacteria are inhibited in their growth by the presence of sulfanilamide or related compounds in the nutrient medium, and it has been suggested by Fildes<sup>1</sup> and by Woods<sup>2</sup> that this inhibition is due to specific interference with the utilization of essential metabolites, in particular *p*-amino benzoic acid. In the present paper data will be presented which show that the growth of isolated tomato roots<sup>3</sup> is inhibited in the presence of appropriate concentrations of sulfanilamide or its derivatives, but that this growth inhibition can be in whole or in part abolished by the further presence of *p*-amino benzoic acid, a substance not otherwise essential as a supplement for the growth of isolated tomato roots.

*Methods.*—Isolated tomato roots were cultured aseptically in the nutrient medium described earlier.<sup>4</sup> One centimeter tips from vigorously growing branch roots were used as inocula for the experimental cultures. Approximately 200 roots were grown in each experiment. The two hundred initial tips were randomly divided into 10 lots of 20 tips each and each treatment carried out on one or more lots. The conclusions contained in this paper are based on a series of 42 such experiments.

All cultures were maintained in Petri dishes, four roots per dish, in the dark at 25°C. for one week. At the end of this time the growth in length of the principal axis of each root was measured. The roots were then discarded and fresh inocula from stock cultures used in further experiments.

Three clones of roots, all of the variety "San Jose Canner"<sup>5</sup> were used. These clones, each of which had previously been subjected to 27 weekly transfers in the standard basal medium, reacted similarly to all of the treatments discussed below.

*Experimental Results.*—Table 1 shows that the growth of isolated tomato roots is inhibited by sulfanilamide and by sulfapyridine. In the presence of 30 mg. per liter of either of these substances isolated tomato roots grew only 15–21% as long in one week as similar roots not receiving inhibitor. Table 1 also shows that the growth inhibition caused by sulfanilamide or sulfapyridine is in part offset by the further addition of *p*-amino benzoic acid to the culture medium in the concentration of 1 mg. per liter. Roots receiving inhibitor and *p*-amino benzoic acid grew 3.3 to 4.5 times as much in one week as roots receiving inhibitor only. This effect cannot be ascribed to a general growth-promoting effect of *p*-amino benzoic acid, since, as shown in table 2, this substance in the concentration of 1 mg. per

liter is without effect on the growth of tomato roots. Increased (10 times) concentrations of nicotinic acid or pyridoxine did not affect the inhibition as did *p*-amino benzoic acid.

TABLE 1  
THE EFFECT OF SULFANILAMIDE AND SULFAPYRIDINE ON THE GROWTH OF ISOLATED TOMATO ROOTS. GROWTH IN MM. PER ROOT PER WEEK

SUPPLEMENTS	SULFONILAMIDE			SULFAPYRIDINE		
	EXPTS.	NO. OF ROOTS	MM./WEEK	EXPTS.	NO. OF ROOTS	MM./WEEK
None	I-25	40	56.8 $\pm$ 2.2	I-24	38	45.0 $\pm$ 2.5
	I-31	26	49.8 $\pm$ 3.1	I-32	32	52.7 $\pm$ 1.6
30 mg. inhibitor/l.	I-25	33	12.0 $\pm$ 1.5	I-24	29	9.7 $\pm$ 1.3
	I-31	14	10.0 $\pm$ 1.3	I-32	30	7.8 $\pm$ 0.6
30 mg. inhibitor/l.	I-25	19	39.5 $\pm$ 3.4	I-24	17	37.9 $\pm$ 2.5
1 mg. <i>p</i> -amino benzoic acid/l.	I-31	16	36.2 $\pm$ 2.3	I-32	14	35.4 $\pm$ 1.4

TABLE 2  
LACK OF EFFECT OF *p*-AMINO BENZOIC ACID ON THE GROWTH OF ISOLATED TOMATO ROOTS. GROWTH IN MM. PER ROOT PER WEEK

EXPTS.	NO <i>p</i> -AMINO BENZOIC ACID		1.0 MG./L. <i>p</i> -AMINO BENZOIC ACID	
	NO. OF ROOTS	LENGTH	NO. OF ROOTS	LENGTH
I-27	18	52.2 $\pm$ 4.3	18	53.6 $\pm$ 3.5
I-31	26	49.8 $\pm$ 3.1	16	49.1 $\pm$ 3.5
I-37	13	42.7 $\pm$ 1.8	17	41.8 $\pm$ 1.3

Sulfathiazole also inhibits the growth of isolated tomato roots as is shown in table 3, which gives the composite results of 16 experiments. Since the growth of untreated control roots varied from one experiment to the next (as in tables 1 and 2) the growth rates in each experiment were reduced to per cent of the untreated control growth rate and it is this relative growth rate which appears in table 3. It is clear from this table that

TABLE 3  
INHIBITION OF THE GROWTH OF ISOLATED TOMATO ROOTS BY SULFATHIAZOLE AND ITS REVERSAL BY *p*-AMINO BENZOIC ACID. MEAN VALUES FROM 16 EXPERIMENTS. GROWTH RATES IN PER CENT OF CONTROL ROOTS GROWN FOR 1 WEEK IN STANDARD BASAL MEDIUM. THE AVERAGE GROWTH RATE (100%) OF THE CONTROL ROOTS WAS 45.8  $\pm$  2.33 MM. PER WEEK IN THESE EXPERIMENTS

SULFATHIAZOLE CONC.; MG./L.	<i>p</i> -AMINO BENZOIC ACID CONCENTRATION; MG./L. GROWTH PER WEEK: % OF UNTREATED CONTROLS				
	0	0.03	0.10	0.30	1.0
100	3	...	...	...	27.5
30	15.2	17.0	21.0	25.0	53.6
10	18.4	20.0	32.3	58.7	94.6
3	37.4	52.2	83.0	99.0	100.7
1	67.6	92.3	97.7	99.0	...
0	100	...	...	...	100

sulfathiazole is a relatively potent inhibitor of root growth since half inhibition is given by about 2 mg. sulfathiazole per liter. The amount of *p*-amino benzoic acid needed to abolish a fixed amount of inhibition varies with the inhibitor concentration. Thus nearly the same absolute increase in growth rate was brought about by 0.03 mg. per liter of the material in the presence of 1 mg. per liter of sulfathiazole as was brought about by 1 mg. per liter of *p*-amino benzoic acid in the presence of 100 mg. per liter of sulfathiazole. Inhibition by a given quantity of sulfathiazole was markedly decreased by the addition of a 100 times smaller quantity of *p*-amino benzoic acid in every case. Inhibition by sulfathiazole and reversal of this inhibition by *p*-amino benzoic acid would seem to depend in part on the ratio between the two substances in the nutrient medium rather than strictly on the absolute amount of inhibitor present.

Isolated tomato roots of the clones used in these experiments appear, qualitatively, to contain *p*-amino benzoic acid or related substance of similar physiological activity,<sup>6</sup> even when grown through 40 or more successive weekly transfers in *p*-amino benzoic acid free medium. The presence of this material in such roots suggests that isolated tomato roots of the present clones may synthesize the substance.

*Discussion.*—The inhibition of the growth of isolated tomato roots by sulfathiazole depends in part on the ratio between added inhibitor and added *p*-amino benzoic acid. This observation might suggest that the inhibition is in part competitive in the sense that sulfathiazole may compete with *p*-amino benzoic acid for some essential position or function in the living cell. That sulfathiazole inhibits the growth of roots not supplied with *p*-amino benzoic acid is not in disagreement with this view since isolated tomato roots appear to normally contain this substance. Other factors may also govern the inhibition.<sup>7</sup>

*Summary.*—Isolated tomato roots of 3 different clones were found to be inhibited in growth by the addition of sulfanilamide, sulfapyridine or sulfathiazole, to the nutrient medium. This inhibition was in whole or in part abolished by the further addition of *p*-amino benzoic acid to the medium. Isolated tomato roots of the clones used normally contain *p*-amino benzoic acid, or a substance having similar physiological activity.

\* Report of work done with the coöperation of the Work Projects Administration, O P. No. 165-1-07-172. This work was made possible in part by the support of Merck and Company.

<sup>1</sup> Fildes, P., *British Jour Expt Path*, 21, 67 (1940).

<sup>2</sup> Woods, D. D., *Ibid.*, 21, 74 (1940).

<sup>3</sup> White, P. R., *Plant Physiology*, 9, 585 (1934). Robbins, W. J., and Bartley, Mary. *Science*, 85, 246 (1937).

<sup>4</sup> Bonner, J., *Amer. Jour. Bot.*, 27, 692 (1940). This medium contains per liter of re-distilled water: 1.5 mg. ferric tartrate, 20 mg.  $\text{KH}_2\text{PO}_4$ , 65 mg. KCl, 81 mg.  $\text{KNO}_3$ , 86 mg.  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 236 mg.  $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ , and 20 gm. sucrose.



<sup>5</sup> The seed used was kindly supplied by the California Packing Corporation.

<sup>6</sup> The assays for *p*-amino benzoic acid were carried out by Prof. G. W. Beadle and Prof. E. Tatum, Stanford University, using a mutant of *Neurospora crassa* for which *p*-amino benzoic acid is an essential supplement.

<sup>7</sup> In 5 experiments sulfanilamide or its derivatives were applied to intact tomato plants grown in the greenhouse in sand or solution culture. Concentrations of up to 125 mg. per liter of nutrient were found to be without marked inhibiting effect on root or top growth.

## GENERAL CONGRUENCES INVOLVING THE BERNOULLI NUMBERS

BY H. S. VANDIVER

DEPARTMENT OF PURE MATHEMATICS, THE UNIVERSITY OF TEXAS

Communicated May 29, 1942

Kummer<sup>1</sup> gave a result which may be expressed as follows:

$$h^n(h^{p-1} - 1)^i \equiv 0 \pmod{p^i}, \quad (n - 1 \geq i; n \not\equiv 0 \pmod{p-1}), \quad (1)$$

where  $p$  is an odd prime; the left-hand member is expanded in full, then  $b_i/t$  is substituted for  $h^i$ , and the  $b$ 's are defined by the recursion formula

$$(b + 1)^n = b_n, \quad (n > 1),$$

in which we expand the left-hand member by the binomial theorem and substitute  $b_k$  for  $b^k$ . The latter formula gives the Bernoulli numbers. The congruences (1) have played an important part in the development of the arithmetic theory of these numbers.

In the present paper we shall employ the theorem which the writer gave in another paper,<sup>2</sup> the statement of which, for convenience, is repeated here:

**THEOREM I.** *Let  $R$  be the ring of algebraic integers in an algebraic field and put*

$$f_{n_i}^{(i)}(\alpha_1, \alpha_2, \dots, \alpha_{k_i}) = \sum_{r=1}^{k_i} \alpha_{ri} a_{ri}^{n_i} \quad (2)$$

where the  $a$ 's and  $\alpha$ 's are in  $R$  and the  $n$ 's are rational integers  $\geq 0$ . Further let there be a rational integer  $d > 0$  such that for all  $r$ 's in the range 1 to  $k_i$  and all  $i$ 's in the range 1 to  $s$ ,

$$a_{ri}^d \equiv 1 \pmod{m} \quad (3)$$

where  $m$  is a fixed ideal in  $R$ . Also, let

$$\beta_1 + \beta_2 + \dots + \beta_s \equiv 0 \pmod{m}$$

where the  $\beta$ 's are in  $R$  and suppose that  $a_i$  is the greatest common ideal divisor of the ideals

$$(a_{ri}), r = 1, 2, \dots, k_i,$$

each of which has a factor,  $\neq (1)$  in common with  $m$ , then

$$(f_{n_1} + f_{n_2} + \dots + f_{n_s})^j \equiv 0 \pmod{(m^j, a_1, a_2, \dots, a_s)} \quad (4)$$

where we expand the left-hand member in full and set

$$f_{n_i}^j = \beta^j f_{n_i + id}^{(i)}(\alpha_1, \alpha_2, \dots, \alpha_{k_i}),$$

$$j = 1, 2, \dots, s.$$

We shall now obtain wide generalizations of (1). In another article the writer defined what was called a Bernoulli number of the  $r$ th order as follows:

$$(m_r b^{(r)} + m_{r-1} b^{(r-1)} + \dots + m_1 b' + m_0)^n \quad (5)$$

$$= b_n(m_r, m_{r-1}, \dots, m_0)$$

$$m_i \neq 0, (i = 1, 2, \dots, r),$$

where the left-hand member is expanded in full and  $b_s$  substituted for

$$b_s^{(k)}; k = 1, 2, \dots, r.$$

For  $r = 1$  and 2, it will be shown that congruences of the type (1) as well as much more general ones also hold for these Bernoulli numbers of higher order. We then use methods which will apply to other formulae involving generalized Bernoulli numbers as well as congruences. They consist in the main of obtaining congruences involving (5) for  $r = 1$  and 2 and then noting that the relations hold for each positive integral value of  $m$ . It is then possible to eliminate the  $m$ 's and obtain entirely new types of formulae involving the ordinary Bernoulli numbers and binomial coefficients only.<sup>3</sup> An example of this depends on the congruence,<sup>4</sup> modulo  $p$ ,

$$t_n(m, k) \equiv \sum_p' \frac{(-1)^{n+1} p^k f_{n-1}(\rho)}{1 - \rho^p} + \sum_{l=0}^{k-1} \sum_p l^{n-1} \rho^{l-k} \quad (6)$$

where

$$\frac{(mb + k)^n - b_n}{n} = t_n(m, k),$$

and

$$f_a(x) \equiv 0^a + x + 2^a x^2 + \dots + (p-1)^a x^{p-1},$$

with

$$0^0 = 1, p \text{ is prime, } (m, p) = 1, m > 1$$

and

$$\sum', \sum_p$$

the first symbol meaning that summation is to be taken over all the distinct  $m$ th roots of unity except unity, and the second symbol meaning summation over all the distinct  $m$ th roots of unity. Each one of the terms in the right-hand members of (6) may be put in the form (2); hence we may apply Theorem I, and obtain, if the  $\beta$ 's are rational integers and  $p$  is a rational prime,

$$\begin{aligned} t_1^{n_1} t_2^{n_2} \dots t_s^{n_s} (\beta_1 t_1^{p-1} + \beta_2 t_2^{p-1} + \dots + \beta_s t_s^{p-1}) & \quad (7) \\ \equiv 0(\text{mod } p^j, p^{n_1-1}, p^{n_2-1}, \dots, p^{n_s-1}) \\ n_i \not\equiv 0(\text{mod } p-1); i = 1, 2, \dots, s, \\ \beta_1 + \beta_2 + \dots + \beta_s \equiv 0(\text{mod } p) \end{aligned}$$

where the left-hand member is expanded in full and  $t_i(m, k)$  substituted for  $t_i^i$ ;  $c = 1, 2, \dots, s$ . In another article<sup>5</sup> I gave a result which may be obtained from (7) by putting  $k = 0$  and noting that we can take  $n$  so that

$$n^{p-1} \equiv 1(\text{mod } p^j). \quad (7a)$$

The relation (7) is only one of a number of relations of this general type which may be obtained from (4); we may, for example, take any of the  $f$ 's in (4) as being unity. Also from (4) we may obtain relations similar to (7) but with the  $\beta$ 's integers in an algebraic field  $R$  and instead of  $p$  we have a prime ideal  $P$  in  $R$ , with also  $(N(P) - 1)$  replacing  $(p - 1)$  in the exponents appearing in the left-hand member of (7), if  $N(P)$  represents the norm of  $P$  in  $R$ .

In particular from (7) we find, for  $n \not\equiv 0(\text{mod } (p - 1))$ ,

$$t^n(t^{p-1} - 1)^j \equiv 0(\text{mod } p^j, p^{n-1})$$

from which we obtain, after employing

$$h^n(h^{p-1} - 1)^j \equiv 0(\text{mod } p^j, p^{n-1}),$$

$$\sum_{i=0}^j (-1)^i \frac{(mb + k)^n + i(p-1)}{n + i(p-1)} \binom{j}{i} \equiv 0(\text{mod } p^j, p^{n-1}) \quad (8)$$

In this we may set  $k = 1$  and let  $m$  belong to the exponent  $r$ , modulo  $p^j$ ,

with  $r \leq p - 1$ , then by setting  $m, m^2, \dots, m^r$  in turn we obtain a set of congruences from which we may eliminate  $m$  and obtain, if  $c < r$ ,

$$\sum_{k=0}^a \sum_{i=0}^j (-1)^i \binom{j}{i} \binom{n + i(p-1)}{c + kr} \frac{b_{c+kr}}{n + i(p-1)} \equiv 0 \pmod{(p^j, p^{n-1})} \quad (9)$$

$$a = \left\lfloor \frac{s-c}{r} \right\rfloor; \quad s = n + i(p-1).$$

This obviously differs from (1) in a number of ways, for example, the subscripts of the  $b$ 's ascend by multiples of  $r$  (a divisor of  $(p-1)$ ) instead of multiples of  $(p-1)$ . Now set  $B_n = (-1)^{n-1} b_n$ ;  $n > 1$  and  $B_0 = 1$ .

In another article<sup>6</sup> the writer discussed the concept of regular primes. First, an improper divisor of a Bernoulli number is defined as a prime number  $p$  such that  $B_n \equiv 0 \pmod{p}$ , with also  $n \equiv 0 \pmod{p}$ . A proper divisor  $p$  of  $B_n$  is such that  $B_n \equiv 0 \pmod{p}$  with  $(n, p) = 1$ . From the relation, if  $n \equiv 0 \pmod{(p-1)}$ ,

$$\frac{b_n}{n} \equiv \frac{b_{n+p-1}}{n+p-1} \pmod{p}$$

it is easy to show that a regular prime is not a proper divisor of any Bernoulli number but an irregular prime is a proper divisor of an infinity of Bernoulli numbers, if we define a regular prime as one that divides at least one of the Bernoulli numbers  $B_n$ ;  $n = 1, 2, 3, \dots, (p-3)/2$ . This provides then a separation of primes into two classes. It is possible to extend all of these notions to numbers of the type  $(mb+k)^n$ . To provide a straight generalization of the notion of regular prime we may first restrict ourselves to  $n$  even. Elsewhere<sup>7</sup> the writer proved that if  $n = n_1 p$ ,  $n_1 \not\equiv 0 \pmod{p-1}$ ,  $(m, p) = 1$ , then  $(mb+k)^n \equiv 0 \pmod{p}$ , where  $m$  and  $k$  are integers,  $p$  prime. In view of this relation we may generalize directly the concepts of regular and irregular primes, and also proper and improper divisors. If none of the numbers  $(mb+k)^{2n}$ ; with  $n = 1, 2, \dots, (p-3)/2$  is divisible by  $p$  we say that  $p$  is  $(m, k)$  regular; otherwise it is termed  $(m, k)$  irregular. In view of the result just quoted, it is easy to show that, employing (8) for  $j = 1$  that a prime which is  $(m, k)$  regular is not a proper divisor of any generalized Bernoulli number of the first order defined by  $m$  and  $k$ , whereas, an  $(m, k)$  irregular prime is the proper divisor of an infinity of  $(m, k)$  Bernoulli numbers. In view of this, corresponding to each integral value of  $m$  excepting  $m = 0$  and each integral value of  $k$ , there is a separation of the rational primes into two classes. We may obviously extend the notion of  $(m, k)$  regular primes to the case where  $n$  is odd and less than  $p-1$ . Primes which do not divide any of the  $(mb+k)^n$ ;  $n = 1, 2, \dots$ ,

$p - 2$ , will be referred to as strongly  $(m, k)$  regular. Such primes exist; it will be found that 5 is a  $(3, 1)$  strongly regular prime and 7 is a strongly  $(3, 2)$  regular prime. Elsewhere<sup>8</sup> the writer proved the congruence theorem concerning ordinary Bernoulli numbers. This theorem also holds if we replace the ordinary Bernoulli numbers by any Bernoulli numbers of the first or second order. Then, by the method indicated in the present paper, we can obtain new relations involving the ordinary Bernoulli numbers. If we employ the relation (5) on page 199 of a paper<sup>4</sup> previously referred to, together with (7a) of the present paper, we obtain, on the elimination of  $m$ , and  $k$ , congruences analogous to (9) but involving products of two Bernoulli numbers in each term.

<sup>1</sup> *Journal für Mathematik*, 41, 368, 372 (1851).

<sup>2</sup> These PROCEEDINGS, 28, 26-27 (1942).

<sup>3</sup> In another article the writer obtained certain equations involving the ordinary Bernoulli numbers by employing initially the generalized Bernoulli numbers. Cf. *Duke Math. Jour.*, 8, 582 (1941).

<sup>4</sup> These PROCEEDINGS, 25, 200 (1939).

<sup>5</sup> *Bull. Amer. Math. Soc.*, 43, 422 (1937).

<sup>6</sup> These PROCEEDINGS, 18, 594-597 (1932).

<sup>7</sup> *Duke Math. Jour.*, 8, 580 (1941).

<sup>8</sup> *Bull. Amer. Math. Soc.*, 46, 121 (1940), Theorem I.

## TRANSFORMATION THEORY OF ISOGONAL TRAJECTORIES OF ISOTHERMAL FAMILIES\*

BY EDWARD KASNER AND JOHN DE CICCIO

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY AND ILLINOIS INSTITUTE OF TECHNOLOGY

Communicated June 23, 1942

1. *Isothermal Transformations.*—It is known that the conformal transformations are the only *point* correspondences of the plane which convert all isothermal families of curves into isothermal families.

Kasner's extension of this result is that the total group of *lineal-element* transformations which send all isothermal families into isothermal families is given in cartesian coördinates  $(x, y, \theta = \arctan dy/dx)$  by

$$X = \phi(x, y), \quad Y = \psi(x, y), \quad \Theta = a\theta + h(x, y), \quad (1)$$

where  $\phi + i\psi$  is a monogenic function of  $z = x + iy$ ,  $h$  is any harmonic function of  $(x, y)$ , and  $a$  is a non-zero constant. The contact part is merely the conformal group.<sup>1</sup>

This is a generalization to transformations of first order differential

elements. In our new work, we shall extend this result to transformations of curvature-elements, that is, second order differential elements.<sup>2</sup>

2. *The Complete System  $\Gamma_0$  of the Isogonal Trajectories of an Isothermal Family.*—To accomplish the purpose stated above, we find that the appropriate geometric object to be studied is the system  $\Gamma_0$  of all the isogonal trajectories of an isothermal family. If  $k$  represents the curvature, the differential equation of any such system  $\Gamma_0$  may be written as

$$k = \lambda(x, y) \sin \theta + \mu(x, y) \cos \theta, \quad (2)$$

where  $\lambda + i\mu$  is a monogenic function of  $z$ .

Any  $\Gamma_0$  system is conformally equivalent to the  $\infty^2$  straight lines of the plane. Thus our system  $\Gamma_0$  may be briefly defined as a *conformal rectilinear wex*.<sup>3</sup>

Various properties of  $\Gamma_0$  systems may be mentioned.<sup>4</sup> The complete set of isogonal trajectories of a family of curves is a linear system if and only if the set is a  $\Gamma_0$  system. The  $\Gamma_0$  systems are the only sets which are both natural and isogonal. The associated point correspondence of a  $\Gamma_0$  system is direct conformal. The reciprocal (conjugate) of a  $\Gamma_0$  system is also a  $\Gamma_0$  system.

3. *Fundamental Theorem.*—The complete group of curvature-element-transformations which send any system  $\Gamma_0$  of all the isogonal trajectories of an isothermal family into another such system  $\Gamma_0$  is given in cartesian coördinates  $(x, y, \theta, k)$  by

$$\begin{aligned} X = \phi(x, y), \quad \pm Y = \psi(x, y), \quad \pm \Theta = \theta + \arctan \frac{\beta(x, y)}{\alpha(x, y)}, \\ K = (\alpha^2 + \beta^2)^{-1/2} [k + \gamma(x, y) \sin \theta + \delta(x, y) \cos \theta], \end{aligned} \quad (3)$$

where  $\phi + i\psi$ ,  $\alpha + i\beta$ ,  $\gamma + i\delta$  are all monogenic functions of  $z = x + iy$  only. (Either both upper or both lower signs are taken.)<sup>5</sup>

A corollary of our theorem is that the only contact transformations of curvature-elements preserving the class of all  $\Gamma_0$  systems are the conformal ones.

4. *Proof of Our Fundamental Theorem.*—For proof, we employ the following minimal coördinates:

$$u = x + iy, \quad v = x - iy, \quad p = \frac{dv}{du} = e^{-2i\theta}, \quad q = \frac{d^2v}{du^2} \frac{dv}{du} = 2ikp^{1/2}. \quad (4)$$

Thus any curvature-element is given by  $(u, v, p, q)$ .

Any  $\Gamma_0$  system is then given by the equation

$$q = c(u) + pd(v). \quad (5)$$

Thus any second order differential equation  $q = q(u, v, p)$  represents a  $\Gamma_0$  system if and only if

$$q_{pp} = 0, q_{up} = 0, pq_{vp} = q_v, q_{uv} = 0. \quad (6)$$

Consider now the general curvature-element transformations

$$U = U(u, v, p, q), V = V(u, v, p, q), P = P(u, v, p, q), Q = Q(u, v, p, q), \quad (7)$$

with non-vanishing jacobian  $J$ . We shall find all such transformations which preserve the class of  $\Gamma_0$  systems.

Let  $q$  defined by (5) be placed into (7). Then  $Q$  becomes a function of  $(U, V, P)$ . We find

$$\frac{\partial Q}{\partial P} = \frac{\alpha + \beta q_u + \gamma q_v + \delta q_p}{a + bq_u + cq_v + dq_p} \quad (8)$$

where

$$a = \begin{vmatrix} U_u U_v U_p \\ V_u V_v V_p \\ P_u P_v P_p \end{vmatrix}, b = \begin{vmatrix} U_u U_v U_p \\ V_u U_v U_p \\ P_u P_v P_p \end{vmatrix}, c = \begin{vmatrix} U_u U_v U_p \\ V_u V_v V_p \\ P_u P_v P_p \end{vmatrix}, d = \begin{vmatrix} U_u U_v U_p \\ V_u V_v V_p \\ P_u P_v P_p \end{vmatrix}, \quad (9)$$

and where  $\alpha, \beta, \gamma, \delta$  are obtained from  $a, b, c, d$ , respectively, by replacing  $P$  by  $Q$ .

Next let  $A$  and  $B$  denote the expressions

$$\begin{aligned} A &= (a\beta - b\alpha) + q_u(c\beta - b\gamma) + q_v(d\beta - b\delta), \\ B &= (a\gamma - c\alpha) + q_u(b\gamma - c\beta) + q_v(d\gamma - c\delta). \end{aligned} \quad (10)$$

Now upon setting  $\partial^2 Q / \partial P^2 = 0$ , and taking account of the conditions  $q_{pp} = q_{up} = 0, q_{vp} = q_v/p$ , we find that for arbitrary  $\Gamma_0$  systems, the partial derivatives  $q_{uu}$  and  $q_{vv}$  will be effectively present in  $\partial^2 Q / \partial P^2$  unless their coefficients vanish. Hence these must be zero, and we find

$$\begin{aligned} A[(U_u V_p - U_p V_u) + q_u(U_v V_p - U_p V_v) + q_v(U_u V_v - U_v V_u)] &= 0, \\ B[(U_u V_p - U_p V_u) + q_u(U_v V_p - U_p V_v) + q_v(U_u V_v - U_v V_u)] &= 0. \end{aligned} \quad (11)$$

Since  $\partial^2 Q / \partial U \partial P = 0$ , there are equations similar to the above where  $U$  is replaced by  $P$ . Also since  $P \partial^2 Q / \partial V \partial P = \partial Q / \partial V$ , there are another set of equations similar to the above where  $V$  is replaced by  $P$ .

We shall prove first that  $A = B = 0$ . For if  $A \neq 0$ , then by (11) and similar equations, it can be shown that if at least one of the quantities  $U_p, U_v, V_p, V_v, P_p, P_v$  is not zero, then  $U_u : U_p : U_v = V_u : V_p : V_v = P_u : P_p : P_v$ . This makes the jacobian vanish. The other possibility,  $U_p = U_v = V_p = V_v = P_p = P_v = 0$ , also makes the jacobian zero. Hence  $A \neq 0$  leads to a contradiction. A similar argument will prove  $B = 0$ .

Next we shall show that  $b = c = \beta = \gamma = 0$ . Since  $A = B = 0$ , it follows from (10) that

$$a\beta - b\alpha = a\gamma - c\alpha = b\gamma - c\beta = b\delta - d\beta = c\delta - d\gamma = 0. \quad (12)$$

Now if at least one of the quantities  $b, c, \beta, \gamma$  is not zero, then  $a:b:c:d = \alpha:\beta:\gamma:\delta$ . This makes the jacobian zero. Therefore  $b = c = \beta = \gamma = 0$ .

Now we shall prove that  $U_p V_q - U_q V_p = 0$ . For otherwise from (9), we discover since  $b = c = 0$  that  $P$  is a function of  $U$  and  $V$  only. This is impossible.

Therefore our required transformations must necessarily satisfy the following equations

$$\begin{aligned} U_p V_q - U_q V_p &= 0, \\ P_p(U_p V_q - U_q V_p) + P_q(U_p V_q - U_q V_p) &= 0, \\ P_p(U_u V_q - U_q V_u) + P_q(U_p V_u - U_u V_p) &= 0, \end{aligned} \quad (13)$$

and similar equations where  $P$  is replaced by  $Q$ .

We shall show next that  $U_q = V_q = 0$ . For otherwise if  $U_q \neq 0$ , it follows from the preceding equations that

$$\begin{aligned} V_q &= \rho U_q, \quad V_p = \rho U_p, \\ (V_u - \rho U_u)(U_p P_q - U_q P_p) &= (V_u - \rho U_u)(U_p P_q - U_q P_p) = 0. \end{aligned} \quad (14)$$

Now if  $U_p P_q - U_q P_p \neq 0$ , then  $V$  is a function of  $U$  only. Since this is obviously impossible, the other possibility is  $U_p P_q - U_q P_p = 0$ . From this, the preceding equations, and similar equations where  $P$  is replaced by  $Q$ , we find that  $U_p:U_q = V_p:V_q = P_p:P_q = Q_p:Q_q$ . This makes the jacobian vanish. Hence  $U_q = 0$ . By a similar argument  $V_q = 0$ .

The equations  $b = c = \beta = \gamma = 0$  now become since  $P_q$  or  $Q_q \neq 0$

$$U_u V_p - U_p V_u = U_u V_p - U_p V_u = 0. \quad (15)$$

If  $U_p$  or  $V_p \neq 0$ , then  $U_u:V_u = U_v:V_v = U_p:V_p$ . This is impossible. Hence in all cases  $U_p = V_p = 0$ .

Therefore our required transformations must necessarily be of the form

$$U = U(u, v), \quad V = V(u, v), \quad P = P(u, v, p, q), \quad Q = Q(u, v, p, q). \quad (16)$$

If this transformation is to convert any  $\Gamma_0$  system into a  $\Gamma_0$  system, then geometrically we observe that, for any fixed point  $(u, v)$  and the corresponding fixed point  $(U, V)$ , a linear equation in  $(p, q)$  must be carried into a linear equation in  $(P, Q)$ . Therefore  $P$  and  $Q$  must necessarily be given by

$$P = \frac{a_2 p + b_2 q + c_2}{a_1 p + b_1 q + c_1}, \quad Q = \frac{a_3 p + b_3 q + c_3}{a_1 p + b_1 q + c_1}, \quad (17)$$

where  $(a_i, b_i, c_i)$  are functions of  $(u, v)$  only.



If  $q = c + dp$  becomes  $Q = C + DP$ , then

$$\begin{aligned} c(b_1C + b_2D - b_3) + (c_1C + c_2D - c_3) &= 0, \\ d(b_1C + b_2D - b_3) + (a_1C + a_2D - a_3) &= 0. \end{aligned} \quad (18)$$

Differentiating the top equation with respect to  $v$ , and the bottom equation with respect to  $u$ , we find

$$\begin{aligned} c \left[ C \frac{\partial b_1}{\partial v} + D \frac{\partial b_2}{\partial v} - \frac{\partial b_3}{\partial v} + b_1 C_v U_v + b_2 D_v V_v \right] \\ + \left[ C \frac{\partial c_1}{\partial v} + D \frac{\partial c_2}{\partial v} - \frac{\partial c_3}{\partial v} + c_1 C_v U_v + c_2 D_v V_v \right] &= 0, \quad (19) \\ d \left[ C \frac{\partial b_1}{\partial u} + D \frac{\partial b_2}{\partial u} - \frac{\partial b_3}{\partial u} + b_1 C_u U_u + b_2 D_u V_u \right] \\ + \left[ C \frac{\partial a_1}{\partial u} + D \frac{\partial a_2}{\partial u} - \frac{\partial a_3}{\partial u} + a_1 C_u U_u + a_2 D_u V_u \right] &= 0. \end{aligned}$$

Since these equations are true for all values of  $C_v$  and  $D_v$ , we obtain

$$\begin{aligned} U_v(cb_1 + c_1) &= 0, & V_v(cb_2 + c_2) &= 0, \\ U_u(db_1 + a_1) &= 0, & V_u(db_2 + a_2) &= 0. \end{aligned} \quad (20)$$

If  $U_u U_v \neq 0$  or  $V_u V_v \neq 0$ , the transformation becomes singular. Hence we have either one of the two cases

$$\begin{aligned} A: & \quad U = \phi(u), & V &= \psi(v); \\ B: & \quad U = \phi(v), & V &= \psi(u). \end{aligned} \quad (21)$$

Substituting case  $A$  into (19) and (20), we find after some manipulation the transformation

$$\begin{aligned} U &= \phi(u), \quad V = \psi(v), \quad P = \frac{\alpha_2(v)p}{\gamma_1(u)}, \\ Q &= \frac{1}{\gamma_1(u)} [q + \alpha_3(v)p + \gamma_3(u)]; \end{aligned} \quad (22)$$

whereas substituting case  $B$  into (19) and (20), we discover

$$\begin{aligned} U &= \phi(v), \quad V = \psi(u), \quad P = \frac{\gamma_2(u)}{\alpha_1(v)p}, \\ Q &= \frac{1}{\alpha_1(v)p} [q + \alpha_3(v)p + \gamma_3(u)]. \end{aligned} \quad (23)$$

Putting these into cartesian coördinates, we finally find our fundamental theorem.

\* Presented to the American Mathematical Society, September, 1942.

<sup>1</sup> Kasner, "Lineal Element Transformations Which Preserve the Isothermal Character," *Proc. Nat. Acad. Sci.*, 27, 406-409 (1941).

Kasner, "Transformation Theory of Isothermal Families and Certain Related Trajectories," *Revista de Matematica*, 2, 17-24 (1941).

De Cicco, "The Two Conformal Covariants of a Field," *Revista de Matematica*, 2, 59-66 (1941).

<sup>2</sup> A generalization of Kasner's theorem has already been given in another direction. This other work is concerned with *field-element* to *lineal-element* transformations. See Kasner and De Cicco, "Generalized Transformation Theory of Isothermal and Dual Families," *Proc. Nat. Acad. Sci.*, 28, 52-55 (1942).

<sup>3</sup> The word *wex* was introduced by Kasner to denote any doubly-infinite system of plane curves. It is therefore represented by a general differential equation of second order  $y'' = f(x, y, y')$ .

<sup>4</sup> See Kasner, *Princeton Colloquium Lectures*, 1912, 1934. Also Kasner, "Isothermal Systems in Dynamics," *Bull. Am. Math. Soc.*, 14, 169-172 (1908); and "A Characteristic Property of Isothermal Systems of Curves," *Mathematische Annalen*, 59, 352-354 (1904).

<sup>5</sup> The content of our new group is expressed by the symbol  $2 \infty^{6f(1)}$ , whereas the content of Kasner's group, equations (1), is  $2 \infty^1 + 4f(1)$ . See Kasner, "A Notation for Infinite Manifolds," *Am. Math. Monthly*, 49, 243-244 (1942).

## DIFFERENTIAL EQUATIONS OF THE TYPE: $y''' = Gy'' + Hy''^2$ \*

BY EDWARD KASNER

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY

Communicated July 1, 1942

1. *Differential Equations of the Type G.*—The purpose of the present article is to discuss the various problems in geometry and especially physics which give rise to the general third order differential equations of the type

$$y''' = G(x, y, y')y'' + H(x, y, y')y''^2. \quad (G)$$

Any such differential equation, or the system of its  $\infty^3$  integral curves will be said to be of *the type G*. In our previous work, we referred to such systems as those possessing *the Property I*.

These differential equations of type *G* were first met in the investigation of dynamical trajectories in general positional fields of force. We shall briefly consider in connection with the differential equations of the type *G* the following topics: (1) The  $\infty^3$  trajectories of a given field of force, (2) the complete system of velocity curves, (3) the systems  $S_k$ , (4) the general catenaries, (5) the  $\infty^3$  brachistochrones in a conservative field, (6) sectional families, (7) curvature trajectories and (8) additive-multiplicative trajectories. All these species of (*G*), except (8), are projectively invariant.

2. *Geometric Properties of Any System of Curves of the Type G.*—Let us

consider any set of  $\infty^3$  curves given by the general differential equation of the third order

$$F: y''' = f(x, y, y', y'').$$

Through any lineal element  $E$ , there pass  $\infty^1$  curves of our family  $F$ . To these curves at  $E$ , construct the osculating parabolas (all those which have four-point contact with the curves at  $E$ ). The following two geometrically equivalent properties, which we call Property I, distinguish all systems of curves of the type  $G$  among all families  $F$ .

( $I_a$ ). *The foci of the  $\infty^1$  osculating parabolas, constructed at the lineal element  $E$ , describe a circle passing through the point of  $E$ .*

( $I_b$ ). *The directrices of these parabolas form a pencil of straight lines.*

In a recent article,<sup>2</sup> Terracini gave the following elegant characterization of the type  $G$  by conic sections. Consider the family  $F$  and let  $C$  be any curve of  $F$  containing the lineal element  $E$ . At  $E$ , construct all the conic sections which have four-point contact with (that is, the same curvature, and the same rate of variation of curvature as)  $C$  at  $E$ . For the single curve  $C$ , there are  $\infty^1$  such conics. Since  $\infty^1$  curves  $C$  of the family  $F$  pass through  $E$ , there are in all  $\infty^2$  conic sections. Terracini's theorem, which is essentially equivalent to our Property I, is:

*The  $\infty^2$  conic sections corresponding to any lineal element  $E$  form a linear family.*

This result illustrates the projective character of the differential equations of the type  $G$ . It may be proved<sup>1</sup> that *the complete group of contact transformations preserving the class of differential equations of the type  $G$  is the projective group consisting of collineations and correlations*. However, the geometry of dynamical trajectories belongs to the group of collineations and not to the total projective group.

3. *Trajectories in an Arbitrary Field of Force.*—Consider the motion of a particle of unit mass in the plane under the action of any positional field of force. The equations of motion are

$$\frac{d^2x}{dt^2} = \phi(x, y), \quad \frac{d^2y}{dt^2} = \psi(x, y), \quad (1)$$

where  $\phi$  and  $\psi$  are the rectangular components of the force acting at any point  $(x, y)$ . We omit the trivial case of zero force.

If a particle is started from any position  $(x_0, y_0)$  with any velocity  $(dx_0/dt_0, dy_0/dt_0)$ , a definite trajectory is described. Since the same curve may be obtained by starting from any one of its  $\infty^1$  points, the total number of trajectories, for all initial conditions, is  $\infty^4$ . By eliminating the time  $t$  from (1), the differential equation of the third order representing this system of trajectories is found to be of the type  $G$  where

$$G = \frac{\psi_x + (\psi_y - \phi_x)y' - \phi_y y'^2}{\psi - y'\phi}, H = -\frac{3\phi}{\psi - y'\phi}. \quad (2)$$

The complete geometric characterization of the  $\infty^3$  trajectories of a given positional field of force has been given in our preceding work (see reference 1).

4. *General Velocity Systems.*—As stated above, if the initial position and the initial velocity of a particle of unit mass are given, the motion is determined. However, if only the initial position and the direction of motion are given, the radius of curvature  $r$  will depend for its value on the initial speed  $v$ . Therefore in addition to the usual formula  $v^2 = (dx/dt)^2 + (dy/dt)^2$ , there must be a formula expressing  $v^2$  in terms of  $x, y, y', r$ . This is furnished by the familiar equation

$$v^2 = rN, \quad (3)$$

where  $N$  denotes the principal normal component of the force, so that

$$N = \frac{\psi - y'\phi}{(1 + y'^2)^{1/2}}. \quad (4)$$

The equation (3) may then be written as

$$v^2 = \frac{(\psi - y'\phi)(1 + y'^2)}{y''}. \quad (5)$$

In the actual trajectory  $v$  varies from point to point. If now we replace  $v^2$  in this result by some constant, say  $1/c$ , the resulting equation may be written as

$$y'' = c(\psi - y'\phi)(1 + y'^2). \quad (6)$$

The curves satisfying this differential equation (they are not in general trajectories) we define as *velocity curves*. For any field, a curve is a velocity curve corresponding to the speed  $v_0$ , provided a particle starting from any lineal element of the curve with that speed describes a trajectory osculating the curve.

In a given field of force, there are  $\infty^3$  trajectories and  $\infty^3$  velocity curves. The complete system of velocity curves is represented by a differential equation of the type  $G$  where

$$G = \frac{\psi_x + y'(\psi_y - \phi_x) - y'^2\phi_y}{\psi - y'\phi}, H = -\frac{\phi}{\psi - y'\phi} + \frac{2y'}{1 + y'^2}. \quad (7)$$

For a given value of  $c$ , there are  $\alpha^3$  velocity curves. The study of such sets of  $\infty^3$  velocity curves is related to *conformal geometry*,<sup>1</sup> while the totality of  $\infty^3$  curves is essentially *projective*.

5. *The Systems  $S_k$ .*—For a given positional field of force, we have thus far discussed the triply-infinite systems of trajectories and velocity curves. Other noteworthy systems of curves are connected with a field, for example brachistochrones, catenaries and tautochrones. Omitting the tautochrones, *the other systems named may all be obtained as special cases of this simple general problem:* to find curves along which a constrained motion is possible such that the pressure is proportional to the normal component of the force.

If an arbitrary curve is drawn in the plane field of force, and a particle of unit mass is started along it from one of its points with a given speed, the constrained motion along the given curve is determined. The acceleration along the curve is given by  $T$ , the tangential component of the force vector. So the speed at any point is

$$v^2 = 2 \int T ds. \quad (8)$$

The pressure  $P$  (of course, normal to the curve) is

$$P = \frac{v^2}{r} - N. \quad (9)$$

*The general problem suggested is to find curves such that  $P$  shall be proportional to  $N$ .* So  $P = kN$ . To a given value of  $k$ , there correspond  $\infty^3$  such curves: the system so obtained will be denoted by  $S_k$ . The four special cases of physical interest are:

- $k = 0$  gives  $S_0$ , the system of *trajectories*;
- $k = -2$  gives  $S_{-2}$ , the system of *brachistochrones*;
- $k = 1$  gives  $S_1$ , the system of *catenaries*;
- $k = \infty$  gives  $S_\infty$ , the system of *velocity curves*.

The differential equation of the system, in intrinsic form, is easily obtained by eliminating  $v$  from the equations

$$v^2 = (k + 1)rN, \quad vv_s = T. \quad (10)$$

The result is

$$Nr_s = {}_nT - rN_s, \quad (11)$$

where  $n = 2/(k + 1)$ .

Placing this into cartesian coördinates, we find that the differential equation of any system  $S_k$  is of the type  $G$  where

$$G = \frac{\psi_x + (\psi_y - \phi_x)y' - \phi_y y'^2}{\psi - y'\phi}, \quad H = -\frac{3\phi}{\psi - y'\phi} - \frac{(n - 2)(\phi + y'\psi)}{(1 + y'^2)(\psi - y'\phi)}. \quad (12)$$

Complete geometric characterizations of the system  $S_k$  have been given (see reference 1).

6. *Sectional Families*.—Now we shall discuss our first purely geometric problem. A *sectional family* of curves is defined by taking all the  $\infty^3$  plane sections of an arbitrary surface and projecting them from some fixed center on to a fixed plane.<sup>4</sup>

Let  $z = \phi(x, y)$  be an arbitrary surface in space. The orthogonal projections of the  $\infty^3$  plane sections of this surface upon the  $xy$ -plane are given in the finite form by

$$ax + by + c\phi(x, y) + d = 0. \quad (13)$$

Eliminating the arbitrary constants  $a, b, c, d$ , we find that the sectional family is given by a differential equation of the type  $G$  where

$$G = \frac{\phi_{xxx} + 3y'\phi_{xxv} + 3y'^2\phi_{xvv} + y'^3\phi_{vvv}}{\phi_{xx} + 2y'\phi_{xv} + y'^2\phi_{vv}}, H = \frac{3\phi_{xv} + 3y'\phi_{vv}}{\phi_{xx} + 2y'\phi_{xv} + y'^2\phi_{vv}}. \quad (14)$$

A complete geometric characterization of any sectional family has been given by Annette Vassel.<sup>5</sup>

7. *Curvature Trajectories*.—To define *curvature trajectories*, we start with an arbitrary doubly infinite family of curves, that is, a general differential equation of second order  $y'' = F(x, y, y')$ . A curvature trajectory of this family is a curve which is drawn so as to have at each point  $c$  times the curvature of the member of the family to which it is tangent at that point,  $c$  remaining constant along the trajectory. For a given value of  $c$  there will be a set of  $\infty^2$  trajectories, one in each direction through each point. By varying  $c$ , we obtain  $\infty^1$  such sets. Hence a given doubly infinite family generates a triply infinite system of curvature trajectories.

The  $\infty^3$  trajectories that can be drawn for a fixed value of the curvature ratio  $c$  satisfy the equation  $y'' = cF(x, y, y')$  (since the curvature at a given lineal element is proportional to  $y''$ ). Eliminating the parameter  $c$  by differentiation, we find that the complete system of  $\infty^3$  trajectories is of the type  $G$  where

$$G = \frac{1}{F} (F_x + y'F_y), H = \frac{F_{y'}}{F}. \quad (15)$$

Such systems have been discussed in relationship to dynamical trajectories and sectional families.<sup>6</sup> We note that *every complete velocity system is a special case of curvature family*.

8. *The Additive-Multiplicative Trajectories*.—We shall first of all consider some elementary lineal element transformations of the plane. These are (1) the *turn* which has been considerably studied in our preceding work,

and (2) the *magnicline*, which has been recently introduced. A turn  $T_\alpha$  is the result of rotating each lineal element about its point through a constant angle  $\alpha$ , whereas a magnicline  $M_m$  is the effect produced by multiplying the inclination  $\theta$  of each lineal element by the same constant number  $m \neq 0$  while leaving its point fixed.

A turn may be considered to be an additive correspondence, while a magnicline is a multiplicate correspondence. The group generated by turns and magniclines is

$$X = x, Y = y, \Theta = a\theta + b, \quad (16)$$

where  $a \neq 0$  and  $b$  are constants. This may be called the additive-multiplicative group. This is intrinsic, independent of axes.

By applying a single transformation of (16) to the lineal elements of a simple family of curves:  $dy/dx = \tan f(x, y)$ , we obtain the  $\infty^1$  additive-multiplicative trajectories of the given family. By varying  $a$  and  $b$ , we obtain the complete system of  $\infty^3$  additive-multiplicative trajectories. Any such system is given by a differential equation of the type  $G$  where

$$G = \frac{f_{xx} + 2y'f_{xy} + y'^2f_{yy}}{f_x + y'f_y}, H = \frac{f_y}{f_x + y'f_y} + \frac{2y'}{1 + y'^2}. \quad (17)$$

Certain generalizations of these trajectories are being investigated by F. McMahon in my seminar.

In conclusion, we may note that all these problems with one single exception belong to the field of projective geometry. The single exception is the additive-multiplicative trajectories, obviously of metric character.

\* Presented to the American Mathematical Society.

<sup>1</sup> Kasner, "Differential Geometric Aspects of Dynamics," *Princeton Colloquium Lectures*, 1912, 1934. Also *Trans. Amer. Math. Soc.*, 1906-1910.

<sup>2</sup> Terracini, "Differential Equations of the Type  $y''' = Gy'' + Hy'^2$ ," *Revista di Matematica*, 2 (1941).

<sup>3</sup> Kasner and De Cicco, "Geometry of Velocity Systems," *Bull. Amer. Math. Soc.*

<sup>4</sup> Kasner, "Dynamical Trajectories and the  $\infty^3$  Plane Sections of a Surface," *Proc. Nat. Acad. Sci.*, 17, 370-376 (1931).

<sup>5</sup> Vassel, "Characterization of Sectional Families," *Am. Jour. Math.* (1939).

<sup>6</sup> Kasner, "Dynamical Trajectories and Curvature Trajectories," *Bull. Am. Math. Soc.*, 449-455 (1934); Comenetz, Curvature Trajectories, *Amer. Jour. Math.*, 58 (1936).

<sup>7</sup> Kasner, "Lineal Element Transformations Which Preserve the Isothermal Character," *Proc. Nat. Acad. Sci.*, 27, 406-412 (1941); also *Revista de Matematicas y Fisica teórica*, II (1941) (Tucumán).

<sup>8</sup> Another important type of differential equation with many applications in calculus of variations and physics is  $y = Ay'''' + By'' + C$ . See Princeton Colloquium and a forthcoming paper by the author in *Revista di Matematica*.

PROCEEDINGS  
OF THE  
NATIONAL ACADEMY OF SCIENCES

Volume 28

September 15, 1942

Number 9

Copyright 1942 by the National Academy of Sciences

OXIDATION-REDUCTION PATTERNS IN AMPHIBIAN AND  
TELEOST DEVELOPMENT

BY C. M. CHILD

SCHOOL OF BIOLOGICAL SCIENCES, STANFORD UNIVERSITY

Communicated July 31, 1942

The problem of amphibian developmental pattern has been attacked by many investigators in widely different ways, but the question of the essential characteristics of this pattern and of vertebrate developmental pattern in general is still under discussion. The following briefly presented data, giving evidence of patterns of enzyme activity, are concerned with this question.

*Methods.*—Intracellular reduction and reoxidation of methylene blue and Janus green and formation, reduction and reoxidation of indophenol in living intact developmental stages of a urodele, *Triturus*, and a teleost, *Oryzias latipes*, show the presence of definite oxidation-reduction patterns, characteristic for particular developmental stages, and undergoing definite changes in the course of development.\* Lightly pigmented and occasionally occurring unpigmented egg masses of *Triturus* permit direct observation of these patterns. *Triturus* embryos remaining within the vitelline membrane after removal of the jelly stain very slowly with oxidized methylene blue, but with decrease of free oxygen in the solution by addition of a minute amount of sodium hydrosulphite the dye, reduced to the colorless "leucobase," penetrates both vitelline membrane and embryo almost at once, and with increase of oxygen recoloration occurs.

The teleost embryos within the chorion stain with oxidized methylene blue, but the chorion stains so rapidly and deeply that it may become difficult to see the embryonic patterns. Oxidized Janus green also stains through the chorion but when reduced to the colorless form by hydrosulphite it penetrates more rapidly and reoxidizes to red.

Intracellular dye reduction is brought about in both amphibian and teleost by oxygen decrease, either rapidly by hydrosulphite, or more slowly by oxygen uptake of embryos sealed in small volume of dye solution. On increase of oxygen after reduction reoxidation of intracellular dye with



recoloration occurs in a definite differential pattern. In the extremely small amounts used, sodium hydrosulphite is not appreciably toxic except perhaps after often repeated additions to the dye solution. Reduction and reoxidation of methylene blue can be repeated several times without affecting further development.

Intracellular formation of indophenol, deep blue in oxidized form, from dimethylparaphenylenediamine (para-aminodimethylaniline) and  $\alpha$ -naphthol, catalyzed by an oxidase, is finally toxic or lethal, but with low concentrations of reagents the intracellular reaction pattern becomes clearly visible before toxic effects appear. The reaction is of little value for intact earlier embryos of *Triturus* because it is extremely slow or inappreciable, but it occurs readily in cells adjoining cut or torn surfaces of isolated pieces in Holtfreter solution or locally injured regions of these stages, and after closure of the neural tube and in young animals after hatching it takes place relatively rapidly. Teleost embryos within the chorion show the reaction at all stages, but more rapidly later than earlier, and much more rapidly in regions or adjoining surfaces where relations of cells have been disturbed by cutting, tearing or isolation of pieces, than in intact parts. After intracellular formation indophenol is reduced to the colorless form and reoxidized with recoloration in the same ways as methylene blue.†

In living amphibian and teleost material in good condition the gradient patterns shown by the indophenol reaction and by reduction of indophenol, Janus green and methylene blue coincide in direction and recoloration on reoxidation progresses in the reverse direction.

*Pattern in Triturus Development.*—In early cleavage reduction of methylene blue and Janus green progresses from the apical (animal) region, more rapidly on one side, presumably the dorsal side. In blastula stages the gradient from the apical region is present and reduction also spreads from an area on the presumably dorsal side. As soon as the earliest stages of invagination permit identification of dorsal and ventral sides with certainty the dorsal lip of the early blastopore appears as the most rapidly reducing region of the embryo and reduction progresses anteriorly and somewhat laterally from it. This dorsal area of rapid reduction is clearly distinguishable to the naked eye and reappears in successive reductions until overstaining or hydrosulphite produces toxic effects. It evidently coincides approximately with the dorsal inductor region but is without distinct boundary, the decrease in reduction toward its anterior and lateral border being gradual. In these early gastrula stages the reduction gradient from the apical region is slight but still visible when pigmentation is not too deep. Dorsally it meets the gradient from the dorsal lip above the equator. The reduction pattern in the dorsal inductor region persists during gastrulation but in later gastrula stages usually appears somewhat less extensive. As the blastopore progresses from crescentic to circular outline,

that is, as invagination extends laterally and ventrally, reduction usually appears somewhat more rapid in lateral and ventral lips, but the gradient does not extend far laterally and ventrally from the lip, and reduction is never as rapid as in the dorsal lip. After closure of the blastopore a small area, chiefly anterior to the blastopore region, still shows relatively rapid reduction for a time.

With beginning of neurulation a further change in pattern appears. Even before the neural plate is distinctly visible reduction becomes more rapid in its anterior region than elsewhere and progresses posteriorly from it. As the neural folds appear they become regions of rapid reduction, also progressing posteriorly. Reduction outside the neural plate is much less rapid and progresses posteriorly and ventrally. In later neurula stages and after closure of the neural tube reduction progresses from the anterior head region posteriorly, rapidly in the dorsal region and posteroventrally and much less rapidly in lateral regions. Early stages of the tail bud are associated with a new region of more rapid reduction and with outgrowth of the bud this becomes a reduction gradient with high end at the tip.

Reduction of methylene blue, Janus green and indophenol and the indophenol reaction become more rapid in early bud stages of the fore leg than in surrounding lateral regions. With outgrowth of the leg the almost radial gradient pattern of the bud becomes longitudinal in consequence of differential growth, with rates of reaction and reduction decreasing from the tip. In the outgrowing leg rates of reaction and reduction also decrease from the anterodorsal to the posteroventral side, that is, the developing leg apparently retains the anterodorsal-posteroventral gradient of the body. Since a longitudinal gradient pattern is also present in it, it possesses pattern in three dimensions long before morphological differentiation is evident and probably from the beginning of its outgrowth.

In outgrowing gill filaments and balancer rates of reduction and of indophenol reaction decrease from the tip. In the gill filament and filament complex of later stages reduction is more rapid on the arterial than on the venous side, probably because of lower oxygen content of blood coming to the gill.

In general, regions which reduce more rapidly show recoloration on reoxidation less rapidly than others, that is, reoxidation gradients are the reverse of reduction gradients, but all gradient patterns may be partially or wholly reversed or obliterated by differential toxic effects of overstaining, indophenol, long continued low oxygen or hydrosulphite.

*Pattern in Oryzias Development.*—With this form intracellular formation, reduction and reoxidation of indophenol were used to a greater extent than reduction and reoxidation of methylene blue and Janus green, but all give the same results. In early cleavage and blastoderm stages the central region of the blastoderm, where cell division is more rapid, seems to show

indophenol reaction and reduction somewhat more rapidly than the margins, but the difference is at best slight. Before invagination begins one side of the blastoderm reacts and reduces more rapidly than other parts. At the earliest stages of invagination it becomes evident that this is the side on which invagination begins, and the region of rapid reduction and reaction corresponds to the region of rapid reduction indicating the dorsal inductor in *Triturus*. In slightly later stages of invagination this dorsal inductor region becomes still more clearly distinguishable with rate of reduction and reaction decreasing anteriorly from the lip of the blastopore and laterally from the prospective median region of the embryo. With development of the germ ring around the yolk, its lateral and ventral regions show slightly more rapid reduction and reaction than regions anterior to them.

As the embryonic area or "shield" forms, reduction and reaction become increasingly rapid in its anterior region where the anterior end of the embryo will develop. At this stage there are two opposed longitudinal gradients in the embryo, one from the dorsal lip, the other from the anterior embryonic region. Rates of reduction and reaction also decrease from the median region laterally. With progress of germ ring over the yolk and closure of the blastopore, the posterior gradient becomes less distinct and disappears, and reduction and reaction progress from the anterior end posteriorly, with a short gradient in the opposite direction as the tail develops. This anteroposterior gradient, together with progress of reduction and reaction from the dorsal region ventrally in head and body, persists in later embryonic stages. The dorsiventral gradient shows the greatest differences in the head region and appears even in the developing eyes. There reduction and indophenol reaction decrease in rate ventrally and slightly posteriorly from the dorsal and slightly anterior region and pigment development follows the same course. Evidently the eye, as a lateral outgrowth, like the leg, retains the general gradient pattern of the body. If this is the case in the urodele eye, it is probably a factor in determining lens regeneration from the dorsal margin of the iris.

In *Oryzias*, as in *Triturus*, cells which have been disturbed in their relations by cutting or tearing the embryo during or after removal from the chorion in a Ringer solution, even in stages before gastrulation, and isolated cell groups and cells show more rapid indophenol reaction than intact embryos or parts, suggesting that oxidase activity may have been increased by the injury. Differences in rate of reduction are more variable because torn or cut regions and isolated cell groups react so much more rapidly than intact embryos or parts that they become deeply colored and may be damaged while reaction is still very slight in the undisturbed cells. With intact embryos and isolated parts in separate solutions there is no certainty as regards intracellular concentrations of indophenol. However,

with more or less similar, apparently non-toxic concentrations, as indicated by color, a far from trustworthy indicator in cell masses of different thickness when viewed by transmitted light, isolated cell groups and cells, and cells adjoining torn or cut surfaces often reduce more rapidly than intact parts and reoxidize more slowly.

*Discussion.*—The indophenol reaction and intracellular reduction and reoxidation of indophenol, methylene blue and Janus green in living intact individuals all show the presence of definite gradient patterns in early development of amphibian and teleost. The more general features of these patterns in the two animals are essentially similar as regards their relations to embryonic axes and prospective regions. The patterns are present long before there is any evidence of morphological differentiation and undergo definite changes in the course of development, with localizations of regions of accelerated reaction and reduction which later induce or develop particular organ systems. All these characteristics indicate that these patterns are expressions of essential physiological factors of development, and that different specific systems of activities and conditions and qualitative differentiations develop from primarily quantitative gradient patterns. The early outgrowing amphibian leg undoubtedly differs specifically in some way from early gill filament and balancer, and all of these differ from the gastrula, but a quantitative gradient pattern appears in all. The local differences characterizing leg, gill filament and balancer originate within, and in definite relation to, the gastrula pattern, and the local differences within leg, gill filament and balancer arise within, and in definite relation to, the gradient patterns of early stages of these organ systems.

Whether regional differences in reduction, reoxidation and indophenol reaction do or do not coincide with regional differences in oxygen uptake and CO<sub>2</sub> production, and whether the assumption is or is not justified that respiration of small isolated fragments of embryos remains the same as when they were parts of the intact individual, it appears evident that respiratory determinations on small isolated fragments do not by any means tell the whole story as regards physiological developmental patterns. Moreover, the indophenol reaction, and less certainly, reduction and reoxidation, indicate that physiological condition may be considerably altered in small isolated pieces of embryos.

\* This study was made possible by the kindness of Dr. V. C. Twitty in making amphibian material available, and of Dr. D. M. Whitaker in permitting use of teleost developmental stages.

† Using low concentrations of reagents, it has been possible to show presence of definite indophenol oxidation-reduction patterns in other living intact organisms. Such patterns in the echinoid, *Dendraster*, have already been described (*Proc. Nat. Acad. Sci.*, 27, 523 (1941)). Data on ciliate Protozoa, and annelid, *Nais paraguayensis*, development of the starfish, *Patiria*, the ascidian, *Clavellina*, and ovaries of *Drosophila* are still unpublished.

*NUCLEOPROTEINS OF CELL NUCLEI*

By A. E. MIRSKY AND A. W. POLLISTER

HOSPITAL OF THE ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH, AND DEPARTMENT  
OF ZOOLOGY, COLUMBIA UNIVERSITY

Communicated August 3, 1942

We have prepared nucleoproteins from a wide variety of animal cells—from mammalian liver, kidney, pancreas, spleen, thymus, brain, from the liver, spleen and blood cells of the dogfish, and from the sperm of the trout, shad, frog and sea urchin. These nucleoproteins are located in the nuclei of the cells from which they are derived. In this paper we shall briefly describe the method of preparation, some properties of the nucleoproteins and the evidence that they are in fact derived from cell nuclei.

*I. Preparation.*—To extract these nucleoproteins from the cell and to separate them from other cellular constituents nothing more drastic is used than neutral sodium chloride solutions of varying concentrations. Before extraction, much cytoplasmic material is removed by thoroughly washing the minced tissue with physiological saline. From liver more than 60 per cent of all the protein present can be removed in this manner without destroying the main outlines of cell structure. The washed tissue is then extracted with 1 *M* NaCl (2 *M* NaCl is needed for extraction of sea-urchin sperm). As soon as the more concentrated salt solution is added the mixture becomes exceedingly viscous. By centrifugation at high speed (10,000 to 12,000 r. p. m.) a viscous, slightly opalescent supernatant fluid is obtained. The supernatant fluid is viscous because of the nucleoprotein dissolved in it. When this solution is added to six volumes of water the nucleoprotein precipitates in a fibrous mass, settling rapidly so that the supernatant fluid can be syphoned off. The precipitate is washed with 0.14 *M* NaCl and then redissolved in 1 *M* NaCl. The solution is centrifuged at high speed to remove any suspended material. The nucleoprotein is reprecipitated by pouring into six volumes of water. If the mixture is stirred with a rod having a crook at its end, the fibrous material generally winds around the rod and adheres to it when the rod is transferred to another vessel (Plate I). The nucleoprotein is again dissolved, centrifuged and precipitated. At this point the preparation is frequently considered to be finished, although purification can be carried further.

For further purification advantage is taken of the unusual solubility of the nucleoprotein. It is soluble in 1 *M* NaCl, insoluble in 0.14 *M* NaCl and soluble again when the salt concentration is reduced to approximately 0.02 *M*. When a solution of nucleoprotein in 1 *M* NaCl is placed in a cellophane tube and dialyzed against water the nucleoprotein first precipitates and then tends to redissolve as the salt concentration within the

cellophane tube continues to drop. The insoluble material is removed by centrifuging. The soluble fraction is precipitated by adding enough NaCl to bring the concentration to 0.14 *M*. After centrifugation the precipitate is dissolved in 1 *M* NaCl. Dialysis can be repeated several times, after which material insoluble in 0.02 *M* NaCl is no longer present.

The quantity of nucleoprotein that dissolves in 0.02 *M* NaCl varies considerably, depending on the source from which it is prepared. Practically all of the nucleoprotein prepared from sheep spleen redissolves when the solution in 1 *M* NaCl is dialyzed. None of the nucleoprotein of trout sperm remains in solution after dialysis against water. Between these two extremes fall the nucleoproteins prepared from other sources.

Once the nucleoprotein has been dissolved in water (or in 0.02 *M* NaCl) there appear changes in its properties; it becomes less viscous, less birefringent when stirred and much less fibrous when precipitated in 0.14 *M* NaCl. These changes persist even after the nucleoprotein is dissolved in 1 *M* NaCl. It may at first appear as if these modifications were the result of a fractionation in which less viscous, birefringent and fibrous material is separated from the bulk of nucleoprotein, but this interpretation appears unlikely because in the case of nucleoprotein prepared from sheep spleen virtually all of the nucleoprotein extracted in 1 *M* NaCl remains in solution when the salt is dialyzed away. In this case, at least, the changes observed are not due to fractionation.

The solubility of the nucleoproteins is probably correlated with the salt concentration of the body fluids. The nucleoprotein from mammals and fresh-water fishes precipitates from solution in 1.0 *M* NaCl when diluted to 0.14 *M* (one part of nucleoprotein solution to six parts of water) a final concentration isotonic with the blood of these animals. In elasmobranch fishes, however, in which the salt concentration of the blood is equivalent osmotically to a 0.50 *M* sodium chloride solution, the fibres appear when the 1.0 *M* solution of nucleoprotein is added to an equal volume of water.

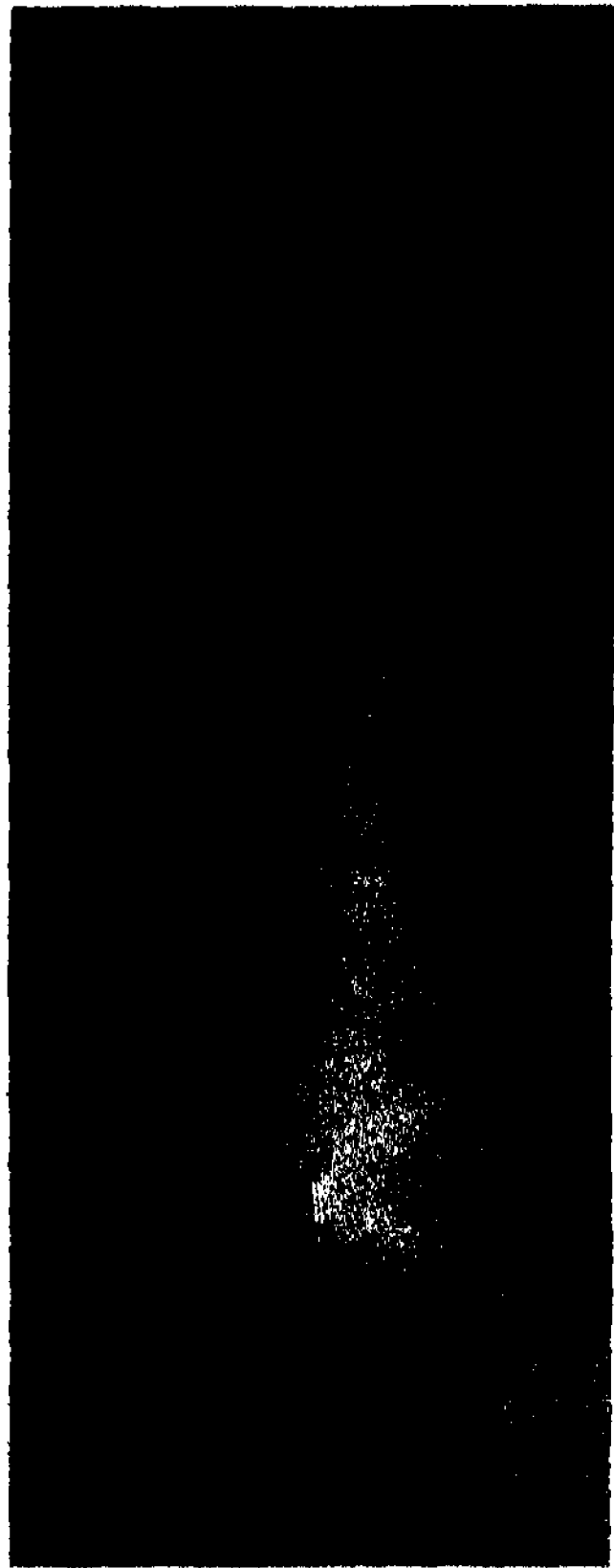


PLATE I



*II. Properties.*—The most striking characteristics of these nucleoproteins are their viscosity and birefringence of flow when in solution and their fibrous nature when precipitated. These properties indicate that the nucleoprotein molecule is markedly elongated.

The phosphorus content of the nucleoprotein of *Arbacia* sperm is 3.05 per cent, of trout sperm 6.55 per cent and of the nucleoproteins prepared from mammalian organs between 3.7 and 4.4 per cent. All of this phosphorus is in the form of desoxyribose nucleic acid. The nucleic acid content of the nucleoproteins ranges from 31 to 66 per cent. From each nucleoprotein the nucleic acid has been isolated and shown to have a composition closely approximating that expected on the tetranucleotide theory. These nucleic acids resemble the highly polymerized nucleic acid prepared from the thymus gland in showing high viscosity and birefringence of flow when in solution and in forming long fibres when precipitated.

The protein component of the nucleoprotein complex has in each instance been prepared. These proteins are histones and protamines. They have a high nitrogen content and basic properties. It is noteworthy that no tryptophane has been found in any of them. Investigation of the SH groups of the histones shows that they are denatured by the method of preparation that has been used in the past.

The ultra-violet absorption spectra of the nucleoproteins show intense absorption at 2540 Å due to the high concentration of nucleic acid present. The protein part of the complex shows in each case (excepting only the protein of trout sperm nucleoprotein, which possesses no aromatic amino acids) a maximum at 2750 Å, slightly removed from the maximum at 2800 Å of a "typical protein" such as egg albumin.

The nucleoproteins prepared from mammalian organs spread at an air-water interface at pH 4, if a little heptyl alcohol is added to the solution. The thickness of the film so formed when measured by the method of Langmuir and Blodgett is 15–16 Å. The protein component alone spreads readily, without the use of heptyl alcohol, at pH 8.3 to give a film 7–9 Å thick, the thickness characteristic of other proteins. The nucleic acid component by itself does not spread at an air-water interface. The spreading experiments on nucleoprotein and on its separated components show that under certain conditions protein and nucleic acid combine to form a complex. Preliminary experiments show that in the Tiselius electrophoresis apparatus, liver nucleoprotein migrates as one electrically homogeneous complex, indicating another condition under which protein and nucleic acid are combined.

There is evidence that the bond between protein and nucleic acid is loose in these nucleoproteins. If a solution (in 1 *M* NaCl) of the nucleoprotein prepared from trout sperm is placed in a cellophane tube and dialyzed against 1 *M* NaCl, the protein (a protamine) gradually diffuses through the

membrane, leaving the nucleic acid behind, so that after prolonged dialysis practically pure nucleic acid remains inside the cellophane tube. Experiments of a different kind with the other nucleoproteins indicate that in them, too, the bond between protein and nucleic acid is loose.

The quantity of nucleoprotein extracted varies considerably. From trout sperm practically all the nucleoprotein present in the cell is extracted but only 20 to 25 per cent of the desoxyribose nucleoprotein present in mammalian organs is extracted.

*III. Site of Origin of Nucleoproteins.*—The preparation of a fibrous protein by methods essentially the same as the one we have used has an exceedingly long history. On numerous occasions since 1865 fibrous proteins have been extracted from mammalian cells and tissues by concentrated solutions of neutral salts and on each occasion the material has been compared with fibrinogen or myosin. The last papers on the subject are by Bensley<sup>1,\*</sup> and by Banga and Szent-Györgyi<sup>2</sup> who, like their predecessors, compare these fibrous proteins with myosin, and believe that they form the structural framework in the cytoplasm of the particular cells from which the protein happened to be prepared. Our investigation of the chemical nature of the fibrous material shows it to be a desoxyribose nucleoprotein and consequently entirely different from myosin. The nucleoproteins we have prepared from various mammalian tissues are essentially the same as the nucleohistone extracted from the thymus gland by somewhat different methods many years ago by Huiskamp<sup>3</sup> and Bang.<sup>4</sup> It is with thymus nucleohistone rather than with myosin that the fibrous materials we have prepared should be compared. Hitherto desoxyribose nucleic acids (those giving the Feulgen reaction) have not been detected by histochemical methods in the cytoplasm; and they have always been found in the cell nucleus. The nucleoproteins of the cytoplasm are widely held to contain only the ribose type of nucleic acid. None of our preparations has been found to contain ribose nucleic acid—not even those extracted from the pancreas, in which direct analyses by previous workers have shown the gland to contain several times as much ribose nucleic acid as desoxyribose.<sup>5</sup> Hence there is a strong presumption that the nucleoproteins that can be extracted by strong saline solutions are located within the cell nuclei. Furthermore, in the lack of tryptophane the protein component differs from most cytoplasmic proteins and resembles the basic protamines of fish sperm and the histones of thymus nucleoprotein.

Both spermatozoa and the lymphocytoid cells of the thymus are types in which the nucleus makes up the bulk of the cell; and the amount of desoxyribose nucleoprotein that can be extracted from these tissues is so large that most of it must have come from the nucleus. For example, the nucleus is approximately nine-tenths of the cell volume in some fish spermatozoa; and over 90 per cent of the dry weight of a suspension of spermatozoa



is extractable as desoxyribose nucleic acid and protamine. But the source of the nucleoprotein cannot be thus determined by calculation alone in extractions from the mammalian liver, in which we have found the nucleus to be somewhat less than one-tenth of the cell volume. For this reason, and especially, moreover, because previous workers, without exception, have believed that the fibrous proteins which they extracted with strong saline solutions came from the cytoplasm of the cells, we have been led to make a direct cytological study of the effects of extraction of the cells.

Observations of individual spermatozoa (of the sea-urchin, *Strongylocentrotus purpuratus*, and the keyhole limpet, *Megathura*) when immersed in the extracting fluid (2.0 *M* NaCl) show that only the nucleus is altered. At first the elongate nucleus swells and becomes spherical; as the swelling continues the outline of the nucleus becomes less definite; and finally the nucleus becomes lost to view. The acrosome, the middle piece with its mitochondrial body and the tail (that is, the entire cytoplasm of the spermatozoon) are not visibly changed; and after the disappearance of the nuclei one finds detached acrosomes and complete tails, with attached middle

#### EXPLANATION OF PLATE II

All figures are photomicrographs of five micra vertical paraffin sections of 100 micra slices (cut on a freezing microtome) of liver of guinea-pig; fixed in Zenker's fluid; and stained by identical schedules with Delafield's hematoxylin and eosin. Each of the larger figures (4, 5, 6, 10, 11, 12) is a highly magnified (1360  $\times$ ) photograph of individual cells from the less magnified (200  $\times$ ) figure immediately above it.

Figures 1 and 4. Normal cells, fixed immediately after sectioning on the freezing microtome; no extraction.

Figures 2 and 5. Frozen section treated with 0.14 *M* NaCl for three hours before being fixed. This has extracted a large part of the stainable material from the cytoplasm. (Negatives and prints of figures 1 and 2, and of 4 and 5 were given identical exposures and were developed simultaneously.)

Figures 3 and 6. Treated with 1.0 *M* NaCl for five minutes before fixation. Nuclei are swollen, no longer stainable, appearing as clear spaces nearly lacking in chromatin.

Figures 7 and 10. Extracted with 1.0 *M* NaCl for five minutes, then placed in 0.14 *M* NaCl (isotonic) for one hour. Compare with figures 3 and 6. The nuclei have shrunk to approximately their original size, and have regained their staining capacity though not the original chromatin pattern.

Figures 8 and 11. Extracted with 1.0 *M* NaCl for one hour. Nuclei swollen and irregular in shape. In many cases nucleus appears to be flowing out of the cell, like a fluid droplet, see lower edge of figure 11.

Figures 9 and 12. Extracted with 1.0 *M* NaCl for one hour, then treated for one hour with 0.14 *M* NaCl. Compare with figures 8 and 11. The isotonic sodium chloride has precipitated the nuclear material (nucleohistone) from its solution in strong saline, in the form of long chromophilic fibres between the cells and along the margin of the section. In many instances these fibres are directly continuous with a mass of similar stainable material in the space originally occupied by the nucleus; for example, see the cell on the left of figure 12.

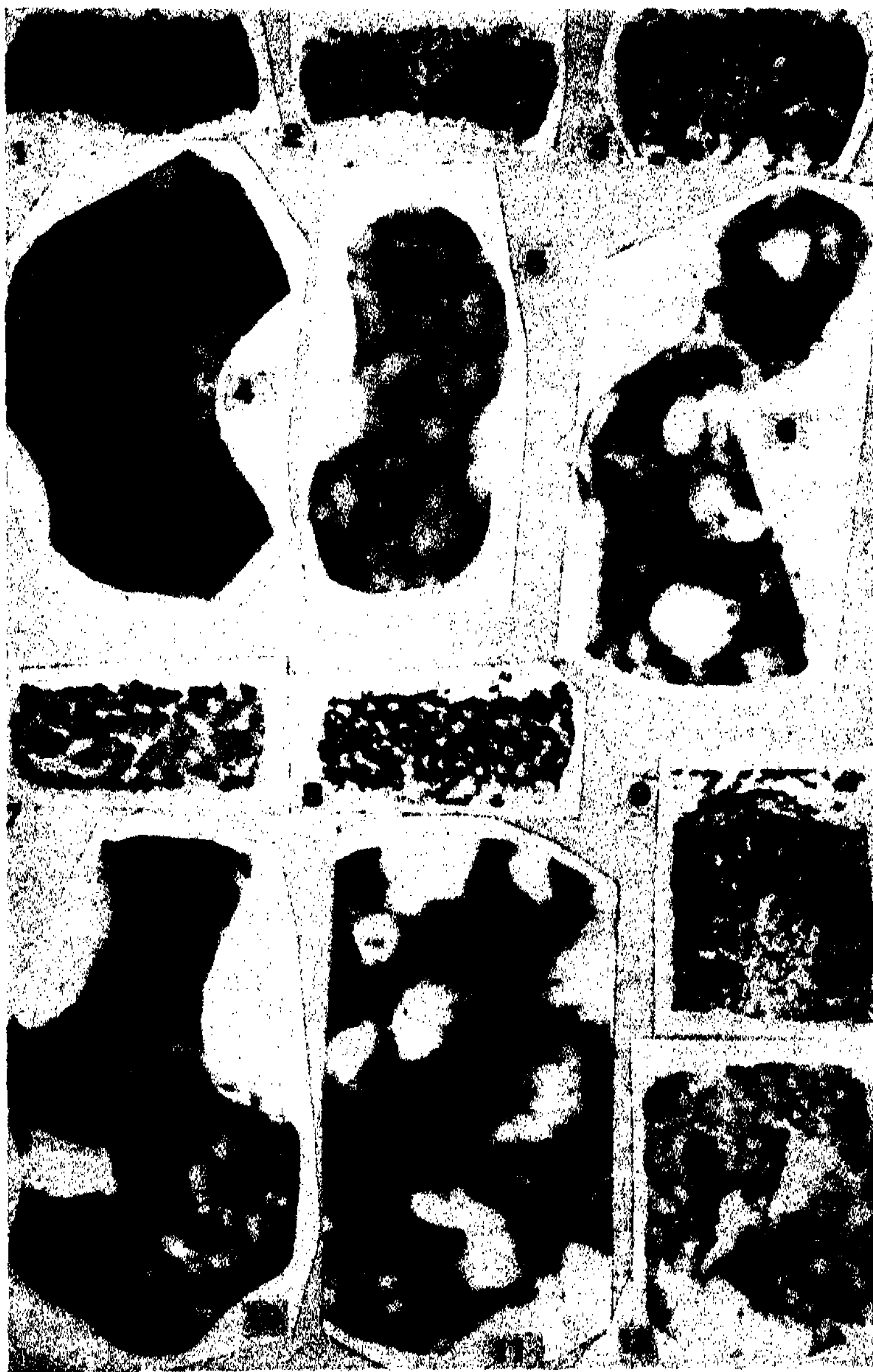


PLATE II

pieces, floating in the salt solution. These cytological observations thus show that strong saline treatment (the method by which one extracts nucleoprotein) causes the nucleus to swell and go into solution. Since no other part of the cell is affected it appears certain that all of the nucleoprotein that can be extracted from sperm suspensions comes from the cell nuclei.

We have studied very thoroughly the effect of extraction of nucleoprotein on the liver cells of the guinea-pig. A few confirmatory observations have been made on other mammalian organs such as pancreas, testis and kidney. For these studies sections of fresh liver, 100 microns thick, were cut on a freezing microtome, and placed in a small volume of extracting fluid for a time. They were then removed and fixed in Zenker's fluid; and thin vertical sections of the 100 microns slice were cut in paraffin. From each experiment one slide was stained progressively with Delafield's hematoxylin and eosin; and a second slide was stained by the Feulgen method. The hematoxylin and the nucleal reaction stained the same material, the chromatin, in the comparable slides. Prolonged treatment of the slice of fresh tissue with 0.14 *M* NaCl (used in the preliminary wash of our method of extraction) removes much of the stainable material from the cytoplasm, but it does not affect the nucleus (compare Plate II, Figs. 2 and 5 with Figs. 1 and 4). Very brief treatment with 1.0 *M* NaCl (the nucleoprotein-extracting fluid), in striking contrast, causes a conspicuous change in the nuclei. They swell slightly, and, surprisingly enough, most of the staining capacity disappears (Plate II, Figs. 3 and 6). This is obviously a swelling preliminary to solution of the nucleoprotein—like the process we have also observed in the spermatozoa (see above). If a slice of tissue in which the nuclei have first been swollen by brief treatment with strong saline is transferred to 0.14 *M* NaCl (the nucleoprotein precipitant) the swollen or dissolved nucleoprotein, as would be expected, is precipitated in the form of irregular fibrous masses that are stainable with hematoxylin and the Feulgen stain (Plate II, Figs. 7 and 10). Although with the precipitation of these fibrous masses the nuclei have reappeared (from the standpoint of stainability) it is noteworthy that none of the recovered nuclei show the uniform, granular pattern of chromatin that characterized them before the treatment (compare Plate II, Figs. 7 and 10 with Figs. 1 and 4). The strong saline causes much more than a mere swelling of the chromatin threads. Quite probably it destroys the intricate arrangement that constitutes what we know as a thread of chromatin, or a chromosome. Long treatment with the strong saline extractant causes the nuclei to become greatly swollen, non-stainable, irregular masses—which often extend outside the cell (Plate II, Figs. 8 and 11). These nuclei quite evidently behave as viscous droplets—with much the same physical properties as those of the freshly extracted solutions of nucleoprotein. The flowing of the nuclei outside the cells—and, indeed, frequently entirely outside the section of

tissue—is visible evidence of what must occur during the extraction of a mass of minced tissue. From these “dissolved” nuclei, in and adjacent to the sections of tissue, the nucleoprotein can be precipitated in the form of long fibres, which are stainable with the nucleal reaction and with hematoxylin (Plate II, Figs. 9 and 12). In these staining capacities the fibres agree with the chromatin of the original nucleus from which they were derived; and it is equally important to note that we have found they share these properties also with the fibres that may be precipitated from solutions prepared by soaking masses of minced tissue in strong saline.

Thus, the cytological observations fully agree with the results of the chemical analysis in showing the nucleus to be the source of the nucleoprotein. From the latter it was evident that the fibrous material extractable with strong sodium chloride solutions is a desoxyribose nucleotide-basic protein complex—a type of compound hitherto found only in cell nuclei. From cytological study of spermatozoa, liver cells, etc., it is clear that the effect of strong sodium chloride solutions is to cause the chromatin of the nucleus to swell and go into solution. Furthermore, from the solution obtained by extraction of minced tissue, nucleoprotein is precipitated by dilution to approximately isotonic strength; and the same treatment of a fresh section of tissue causes precipitation of a fibrous mass within the viscous droplet that each nucleus becomes when its nucleoprotein is dissolved in strong saline solution.

*IV. Summary and Conclusion.*—A method is described for extraction of nucleoproteins from cell nuclei of spermatozoa, and of a variety of other tissue cells. The nucleoprotein is a highly polymerized desoxyribose tetranucleotide in loose combination with basic protein (protamine or histone). This nucleoprotein of cell nuclei is thus found to have the composition suggested by the early and classical studies of Miescher and of Kossel of the separate and degraded components of the nucleoproteins of spermatozoa, prepared by much more drastic and destructive methods.† The nucleoprotein extracted in the present study is in the form of relatively large molecules, or particles; and it is therefore probably highly polymerized. We suggest that in this form it is much closer to its actual condition in the cell nucleus than when in the form in which it has been extracted by other methods. The wide applicability of the present method is a unique advantage, and one that is most promising; for it makes it appear possible, for the first time, to project a study of relatively intact nucleoprotein complexes on a wide scale.

In the course of this work we have been fortunate in having the collaboration of Dr. George Lavin and Dr. Alexandre Rothen of the Rockefeller Institute and of Mr. Stanley Walters of the New York State Fish Hatchery at Cold Spring Harbor.

\* This extraordinarily interesting paper contains a valuable list of references.

† Kossel believed that basic proteins "do not, however, occur in all nuclei, but only in the nuclei of certain kinds of tissues."<sup>6</sup>

<sup>1</sup> Bensley, R. R., *Anat. Rec.*, 72, 351 (1938).

<sup>2</sup> Banga, I., and Szent-Györgyi, A., *Enzymologia*, 9, 111 (1941).

<sup>3</sup> Huiskamp, W., *Zeit Physiol. Chem.*, 32, 145 (1901).

<sup>4</sup> Bang, I., *Beitr. Chem. Phys. Path.*, 4, 115, 331 (1903).

<sup>5</sup> Jorpes, E., *Acta Med. Scand.*, 68, 253, 503 (1928).

<sup>6</sup> Kossel, A., *The Protamines and Histones*, Longmans, Green and Company, 1928, page vii.

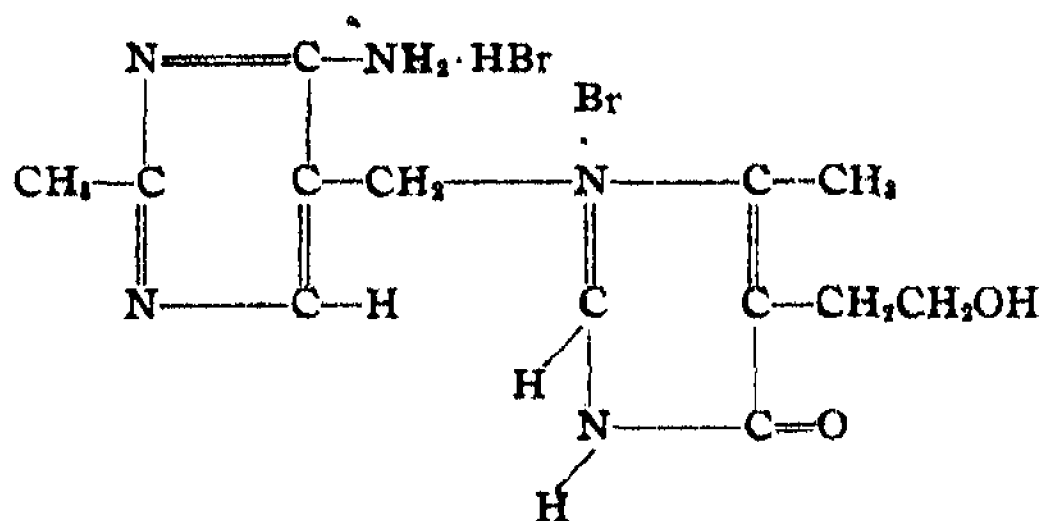
## A PYRIMIDINE ANALOG OF THIAMINE AND THE GROWTH OF FUNGI

BY WILLIAM J. ROBBINS

NEW YORK BOTANICAL GARDEN AND DEPARTMENT OF BOTANY, COLUMBIA UNIVERSITY

Communicated August 7, 1942

Through the courtesy of Dr. R. C. Elderfield I received a pyrimidine analog of thiamine having the following formula:



This compound was synthesized by Miss Yolanda A. Tota. It is stated to be at least as stable to heat as thiamine.

I have tested this compound with three fungi, *Phycomyces Blakesleeanus*, *Pythiomorpha gonapodyides* and *Phytophthora cinnamomi*, each of which has a different kind of thiamine deficiency.

Each organism was grown in 25-ml. quantities of a solution containing per liter 50 g. dextrose, 1.5 g.  $\text{KH}_2\text{PO}_4$ , 0.5 g.  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ , 2.0 g. asparagine and the following mineral supplements in p. p. m.: 0.005 B, 0.02 Cu, 0.1 Fe, 0.01 Ga, 0.01 Mn, 0.01 Mo and 0.09 Zn. To this solution various amounts of the analog, of thiamine, of thiazole or of pyrimidine<sup>1</sup> were added as shown in tables 1 and 2. The solutions were sterilized for 20 minutes at 12 pounds pressure and each treatment was carried out in triplicate. The temperature of incubation was 20°C. *Phycomyces* was grown for 8 days,

*Phytophthora* for 11 days and *Pythiomorpha* for 11 days in the first experiment, table 1, and 15 days in the second experiment, table 2. At the end of the period of growth the mycelium was removed from each flask, washed with distilled water, dried at 100°C. and weighed.

The pyrimidine analog was neither beneficial nor detrimental to *Phycomyces* even when 1000.0 m $\mu$  moles were added per flask (tables 1 and 2). The addition of pyrimidine to the basal solution containing 2.0 m $\mu$  moles of the analog did not improve growth. These results are to be expected since all previous evidence shows that both pyrimidine and thiazole must be available for growth of *Phycomyces* to occur. Addition of thiazole to 2.0 m $\mu$  moles of the analog had little effect as may be noted by comparing

TABLE 1  
DRY WEIGHT OF MYCELIUM PRODUCED IN SOLUTION OF MINERALS, ASPARAGINE AND DEXTROSE SUPPLEMENTED AS INDICATED

ADDITIONS PER FLASK CONTAINING 25 ML. OF THE BASAL SOLUTION	AV. DRY WT. PER CULTURE MG.		
	PHYCOMYCES	PYTHIOMORPHA	PHYTOPHTHORA
None	0.5	1.2	3.8
0.5 m $\mu$ mole analog	0.8	1.5	2.2
1.0 m $\mu$ mole analog	0.1	2.1	0.2
2.0 m $\mu$ moles analog	0.2	2.5	2.8
10.0 m $\mu$ moles analog	0.7	5.8	2.5
100.0 m $\mu$ moles analog	0.4	20.0	0.4
1000.0 m $\mu$ moles analog	2.9	84.5	0.8
2.0 m $\mu$ moles analog + 100.0 m $\mu$ moles pyrimidine	1.6	113.9	0.6
2.0 m $\mu$ moles analog + 100.0 m $\mu$ moles thiazole	4.3	2.9	0.6
0.5 m $\mu$ mole thiamine	77.8	32.9	93.9
1.0 m $\mu$ mole thiamine	138.3	61.5	162.4
2.0 m $\mu$ moles thiamine	160.5	74.7	150.1

the growth obtained with that combination and in the solution containing 2.0 m $\mu$  moles of thiamine (table 1). However, when larger quantities of the analog (10.0, 100.0 or 1000.0 m $\mu$  moles) were used in the presence of 100 m $\mu$  moles of thiazole considerable growth developed (table 2). The dry weight with 100.0 m $\mu$  moles of the analog in the presence of 100.0 m $\mu$  moles of thiazole was about the same as that found with 0.1 m $\mu$  mole of pyrimidine and 100.0 m $\mu$  moles of thiazole; 1000.0 m $\mu$  moles of the analog in the presence of thiazole produced more dry weight than 1.0 m $\mu$  mole of pyrimidine and 100.0 m $\mu$  moles of thiazole. The analog was ineffective as a substitute for thiamine for *Phycomyces* but it was about  $1/1000$  as effective as thiamine as a source of pyrimidine.

*Pythiomorpha* grows if furnished with the pyrimidine half of the thiamine molecule; it is able to synthesize the thiazole portion from the sugar, minerals and asparagine in the basal medium. The analog was not in-

jurious to *Pythiomorpha*, in fact 10.0 m $\mu$  moles definitely improved its growth and 100.0 or 1000.0 m $\mu$  moles were still more favorable (tables 1 and 2). However, less growth was obtained with 100.0 m $\mu$  moles of the analog than with 0.5 m $\mu$  mole of thiamine. In the first experiment somewhat more growth occurred with 1000.0 m $\mu$  moles of the analog than with 2.0 m $\mu$  moles of thiamine (table 1). In the second experiment 1000.0 m $\mu$  moles gave about the same yield as 2.0 m $\mu$  moles of thiamine and somewhat more than 0.5 m $\mu$  mole of thiamine or of pyrimidine (table 2). As a source of pyrimidine for *Pythiomorpha* the analog appeared to be about  $1/500$  as effective as thiamine.

TABLE 2

DRY WEIGHT OF MYCELIUM PRODUCED IN SOLUTION OF MINERALS, ASPARAGINE AND DEXTROSE SUPPLEMENTED AS INDICATED

ADDITIONS PER FLASK CONTAINING 25 ML. OF THE BASAL SOLUTION	AV. DRY WT. PER CULTURE MG.	
	PHYCOMYCES	PYTHIOMORPHA
None	Trace	3.7
10.0 m $\mu$ moles analog	Trace	7.6
100.0 m $\mu$ moles analog	Trace	30.7
1000.0 m $\mu$ moles analog	Trace	100.1
10.0 m $\mu$ moles analog + 100 m $\mu$ moles thiazole	9.9	6.0
100.0 m $\mu$ moles analog + 100 m $\mu$ moles thiazole	33.9	39.4
1000.0 m $\mu$ moles analog + 100 m $\mu$ moles thiazole	136.2	96.0
100.0 m $\mu$ moles thiazole	1.4	0.3
0.5 m $\mu$ mole thiamine	91.0	56.3
1.0 m $\mu$ mole thiamine	127.1	79.7
2.0 m $\mu$ moles thiamine	142.8	102.4
0.1 m $\mu$ mole pyrimidine + 100 m $\mu$ moles thiazole	33.2	62.9
0.5 m $\mu$ mole pyrimidine + 100 m $\mu$ moles thiazole	115.5	79.3

*Phytophthora* will not grow unless supplied with molecular thiamine. The analog, even in amounts of 1000.0 m $\mu$  moles per flask was ineffective with this organism (table 1).

The response of the three fungi leads to the following conclusions: The pyrimidine analog of thiamine does not replace thiamine in the physiology of these organisms. Its effectiveness as a source of pyrimidine for *Phycomyces* or *Pythiomorpha* is of the order of  $1/500$  or  $1/1000$  that of thiamine. The action of the analog as a source of pyrimidine might be ascribed to the presence of traces of pyrimidine as an impurity or more probably to a slight dissociation of the compound into its constituents or to decomposition in sterilization. Neither *Phycomyces* nor *Pythiomorpha* appears able to split



the compound and obtain pyrimidine from it. In this respect the pyrimidine analog differs from the pyridine analog studied earlier;<sup>2</sup> the latter compound was about as effective a source of pyrimidine as thiamine.

<sup>1</sup> The terms pyrimidine and thiazole are used in this paper to refer to the intermediates of thiamine, 4-methyl-5- $\beta$ -hydroxyethyl thiazole and 2-methyl-5-bromomethyl-6-aminopyrimidine hydrobromide.

<sup>2</sup> Robbins, W. J., *Proc. Nat. Acad. Sci.* 27, 419 (1941).

## ON THE PHYSICAL CHARACTERISTICS OF THE PERSEUS CLUSTER OF NEBULAE

BY F. ZWICKY

NORMAN BRIDGE LABORATORY OF PHYSICS, CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated July 31, 1942

*A. The Radial Distribution of Nebulae in the Perseus Cluster.*—In a previous paper<sup>1</sup> counts of nebulae in the field which covers the Perseus cluster were communicated. These counts refer only to nebulae which can be distinguished on limiting exposures taken with the 18-inch Schmidt telescope on Palomar Mountain. As was done previously<sup>2</sup> with the data on the clusters in Coma and in Hydra we here compare the distribution of the nebulae in the Perseus cluster with the distribution which Emden<sup>3, 4</sup> deduced theoretically for the bounded isothermal gravitational gas sphere. In table 1 are listed the average numbers  $N_r$  per square degree of the nebulae brighter than about the absolute photographic magnitude  $M_p = -14.3$  as a function of the distance  $r$  from the center of the Perseus cluster. From the numbers  $N_r$  previously<sup>1</sup> given, the numbers  $N_r$  listed here are obtained by subtracting 3.0 nebulae per square degree which represent the average background of the field nebulae in which the Perseus cluster appears imbedded. The reduction of the observed distribution to the standard Emden curve of the projected densities  $D$  of the isothermal gas sphere is accomplished by plotting in figure 1 the values of  $5.5N_r$  as a function of the associated Emden radius  $r_1$  which is related to the actual radius vector by the equation

$$r = \alpha r_1. \quad (1)$$

A good fit between the observational data and the theoretical curve is obtained if the structural length (or structural index)<sup>5</sup>  $\alpha$  is set equal to 10/3 minutes of arc or, in absolute measure, if we take 11 million parsecs as the distance of the Perseus cluster

$$\alpha = 3.30 \times 10^{21} \text{ cm.} \quad (2)$$



In figure 1 the drawn-out curve represents the radial distribution<sup>4</sup> of the projected density  $D - 37/1000$  of that bounded isothermal Emden sphere which is obtained from the density  $D$  of the infinite sphere by subtracting the constant  $37/1000$ . Although this constant must be expected to vary from cluster to cluster it is seen from figure 1 that the adoption of the value

TABLE 1  
RADIAL DISTRIBUTION OF NEBULAR IN THE PERSEUS CLUSTER

$r$ IN MINUTES OF ARC	$N_r$ PER SQUARE DEGREE	$5.5N_r$	$r_1$	$1000D$	$1000D - 37$
5	554	3047	1.5	2430	2393
10	214	1177	3	1524	1487
15	145	798	4.5	968	931
20	96	528	6	649	612
25	59	325	7.5	476	439
30	52	286	9	366	329
35	51	281	10.5	305	268
40	34	187	12	254	217
45	22	121	13.5	..	..
50	21.4	117	15	195	158
60	16.0	88	18	157	120
70	15.7	86.4	21	..	..
80	10.8	59.4	24	..	..
90	12.0	66.0	27	..	..
100	10.4	57.2	30	92.0	55
110	5.3	29.2	33	..	..
120	8.0	44.0	36	..	..
130	7.7	42.4	39	..	..
140	5.2	28.6	42	..	..
150	4.6	25.3	45	63.1	26.1
160	4.1	22.6	48	..	..
170	1.22	6.7	51	..	..
180	2.30	12.7	54	..	..
190	2.33	12.8	57	..	..
200	1.27	7.0	60	49	12.0
210	1.40	7.7	63	..	..
220	2.93	16.1	66	..	..
230	5.25	28.9	69	..	..
240	0.95	5.2	72	..	..
250	1.73	9.4	75	39	2.0

$37/1000$  used already in the analysis of the clusters in Coma and in Hydra results in an equally satisfactory representation of the nebular counts in the Perseus cluster. Because of the logarithmic scale used in figure 1 for  $N_r$ , the fluctuations, which are proportional to  $(N_r/r)^{1/2}$ , appear inordinately large for small values of  $N_r$ . The absolute values of the fluctuations are, however, within the theoretically expected range. Like the clusters in Coma and in Hydra, the Perseus cluster may consequently be considered as

a large scale assembly of nebulae which is statistically stationary. Basing our further considerations on this result we may again, as in the case of the clusters in Coma and in Hydra, carry through the quantitative analysis described elsewhere.<sup>2</sup> Proceeding in this manner we expect to arrive at a correct prediction for the magnitude of the velocity dispersion in the Perseus cluster in dependence of its structural length  $\alpha$  and its central density  $\rho_0$ .

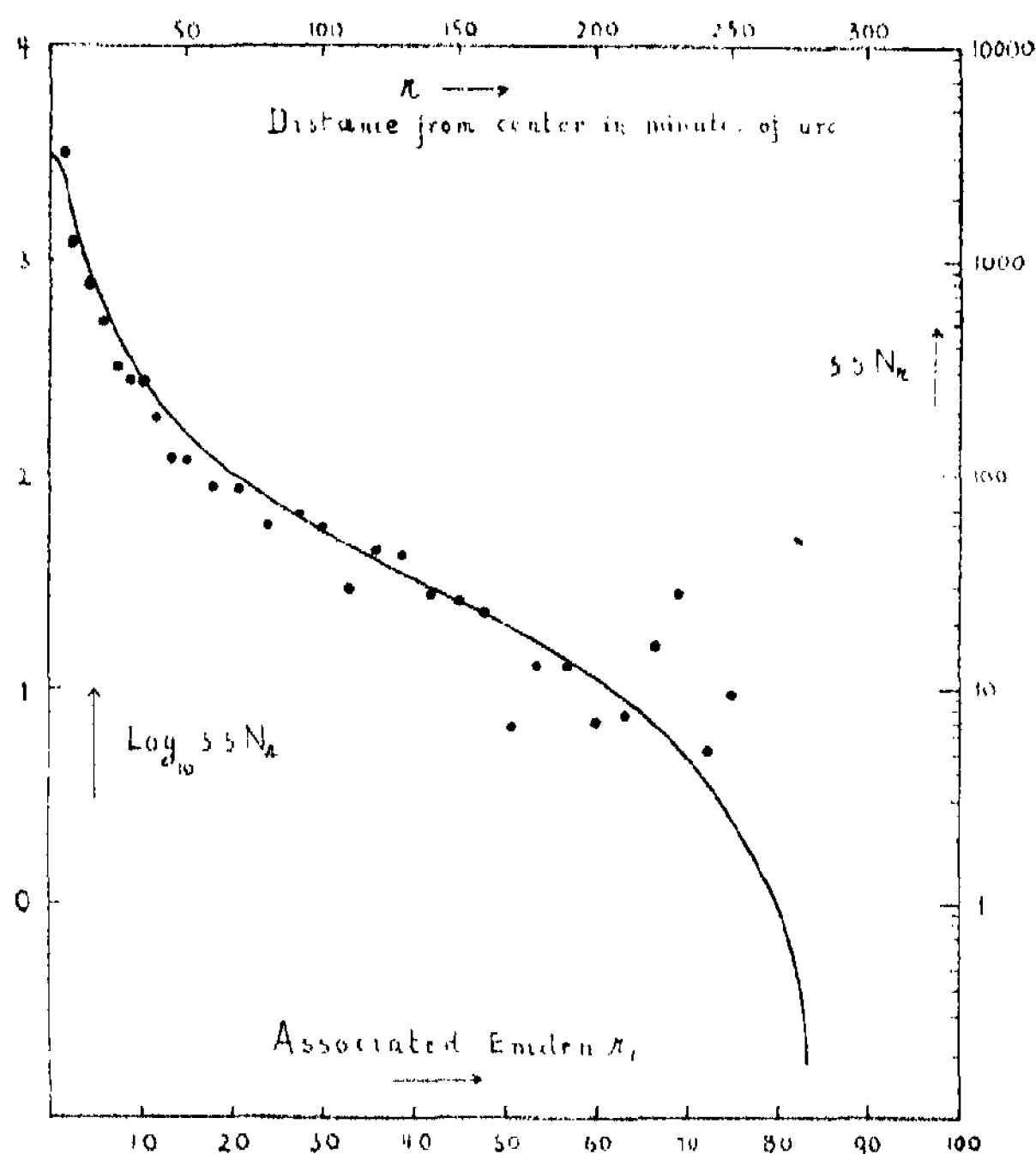


FIGURE 1

*B. Some Physical Characteristics of the Perseus Cluster.*—In the Emden isothermal sphere the projected central density  $\sigma_0$  and the real space density  $\rho_0$  are connected by the equation<sup>2</sup>

$$\rho_0 = \sigma_0 / 3.03\alpha. \quad (3)$$

Choosing the megaparsec as the unit of length, introducing  $\alpha = 10/3$  minutes of arc  $= 1/18$  degree and  $\sigma_0 = 3030/5.5$  nebulae per square degree, and finally converting angular measures into megaparsecs we obtain for the center of the cluster the density

$$\rho_0 = 460,000 \text{ nebulae per cubic megaparsec.} \quad (4)$$

This number includes only the nebulae listed in table 1 which are all brighter

than about the absolute photographic magnitude  $M_p = -14.3$ . We denote with  $\bar{M}$  the average mass of these nebulae and we obtain for the real average central density of matter per cm.<sup>3</sup>

$$\rho_0 = 4.6 \times 10^5 \gamma \bar{M} / 3 \times 10^{13} = 1.53 \times 10^{-27} \gamma \bar{M}, \quad (5)$$

where  $\gamma$  is of the order of unity when it is assumed that the actual central density of matter is not materially greater than that incorporated in the brighter nebulae here considered. As was shown before<sup>5</sup> the velocity dispersion  $(\bar{w}^2)^{1/2}$  can be calculated from the relation

$$(\bar{w}^2)^{1/2} = \alpha [12\pi\Gamma\rho_0]^{1/2} \quad (6)$$

where  $\Gamma$  is the universal gravitational constant. The following table 2 gives  $\rho_0$  and  $\gamma\bar{M}$  in dependence of various assumed values for the velocity dispersion  $(\bar{w}^2)^{1/2}$ . The mass of the sun is  $M_\odot = 2 \times 10^{33}$  g.

TABLE 2

CENTRAL DENSITY  $\rho_0$  AND AVERAGE MASS  $\bar{M}$  OF THE NEBULAE BRIGHTER THAN  $M_p = -14.3$  IN THE PERSEUS CLUSTER IN DEPENDENCE OF THE VELOCITY DISPERSION

$(\bar{w}^2)^{1/2}$ IN KM./SEC.	$\rho_0$ IN G./CM. <sup>3</sup>	$\gamma\bar{M}/M_\odot$
250	$2.2 \times 10^{-25}$	$7.1 \times 10^9$
500	$8.8 \times 10^{-26}$	$2.8 \times 10^{10}$
750	$2.0 \times 10^{-24}$	$6.4 \times 10^{10}$
1000	$3.5 \times 10^{-24}$	$1.1 \times 10^{11}$
1250	$5.5 \times 10^{-24}$	$1.8 \times 10^{11}$
1500	$7.9 \times 10^{-24}$	$2.6 \times 10^{11}$

According to information kindly supplied me by Dr. Hubble the data on the radial velocities  $w_r$  of nebulae in the Perseus cluster are still very scant. An estimate on the basis of these data indicates that  $(\bar{w}_r^2)^{1/2} \cong 350$ –400 km./sec. which results in

$$(\bar{w}^2)^{1/2} = (3\bar{w}_r^2)^{1/2} \cong 600$$
–700 km./sec. (7)

Unless faint nebulae or dark matter contribute far more to the mass of the Perseus cluster than the nebulae which are brighter than  $M_p = -14.3$  the average mass of these nebulae becomes of the order of  $5 \times 10^{10} M_\odot$ . Although we cannot at the present time check this conclusion directly we may apply the following quantitative test to the interpretation given here of the observations on the Perseus cluster by comparing these observations with those made on the Coma cluster. If we denote with the indices  $P$  and  $C$  the quantities which refer to the clusters in Perseus and in Coma, respectively, we have according to (6)

$$(\bar{w}^2)_P = [\alpha_P^2 \rho_{0P} / \alpha_C^2 \rho_{0C}] (\bar{w}^2)_C. \quad (8)$$

Inserting the numerical values for  $\alpha_P/\alpha_C$  and  $\rho_{0P}/\rho_{0C}$  it follows that

$$(\bar{w}^2)^{1/2}_{\text{Persei}} < 0.62(\bar{w}^2)^{1/2}_{\text{Coma}}, \quad (9)$$

where the "smaller" sign means "somewhat smaller" since the ratio  $\rho_{0P}/\rho_{0C}$  was set equal to the ratio per unit cube of the nebulae brighter than the absolute magnitudes  $M_p = -14.3$  and  $M_p = -14.5$  for the clusters in Perseus and in Coma, respectively. Adopting 1100 km./sec. for the dispersion in radial velocities in the Coma cluster this leads to

$$(\bar{w}^2)^{1/2}_{\text{Persei}} = \text{somewhat smaller than } 680 \text{ km./sec.} \quad (10)$$

for the predicted dispersion in the radial velocities of the nebulae in the Perseus cluster, a result which compares favorably with the dispersion of 400 km./sec. derived by Hubble as a very rough estimate from a few directly measured velocities.

TABLE 3  
PHYSICAL CHARACTERISTICS OF THE CLUSTERS OF NEBULAE IN  
COMA HYDRA PERSEUS

Distance in $10^6$ light years	45.0	23.8	35.9
Diameter in $10^6$ light years	4.4	4.7	6.0
Diameter in minutes of arc	340	680	566
Total number of nebulae brighter than photog. mag. $m_L$	670	270	360
Limiting photog. mag. $m_L$	16.6	16.2	16.5
$m_{\text{max.}}$ = photog. mag. of brightest nebula	14.1	13.1	13.8
$M_{\text{max.}}$ = estimated absolute photog. mag. of brightest nebula	-17.0	-17.0	-17.0
$M_L = M_{\text{max.}} + (m_L - m_{\text{max.}})$	-14.5	-13.9*	-14.3
$\alpha$ = structural index	$2' =$ $2.48 \times 10^{22} \text{ cm.}$	$4' =$ $2.56 \times 10^{22} \text{ cm.}$	$3.33' =$ $3.30 \times 10^{22} \text{ cm.}$
Number of nebulae per cubic megaparsec in center of cluster with $M \leq M_L$	2,100,000	880,000	460,000

\* In a previous paper<sup>2</sup>  $M_L$  for the Hydra cluster by mistake was given as  $-13.1$  instead of  $-13.7$  as would have followed from the considerations there used.

*C. Review of Some of the Physical Characteristics of the Clusters of Nebulae in Coma, in Hydra and in Perseus.*—The clusters in Coma, in Hydra and in Perseus are the three largest among the spherically symmetrical clusters which are in reach of the 18-inch Schmidt telescope on Palomar Mountain. It was shown in a series of investigations<sup>2</sup> that the observed radial distribution curves of the brighter nebulae in these clusters can all be reduced to the

same standard curve which represents the radial distribution of the density of matter in a bounded isothermal gravitational gas sphere. The two reduction factors involved in the reduction of the observational data to the theoretical curve can be expressed in terms of the velocity dispersion in the cluster, its central density or the average mass and number per unit volume of the nebulae involved. The substitution of observed values for the quantities mentioned into the theoretical relations results in a satisfactory check of these relations and furnishes significant support for the following contentions: (1) clusters of nebulae represent statistically stationary distributions of matter as far as the brighter nebulae are concerned; (2) Newton's law of gravitation as first good approximation satisfactorily represents the interactions of nebulae separated by distances of the order of one million light years; and (3) the masses of the brighter nebulae are of the order of  $10^{10} M_{\odot}$  to  $10^{11} M_{\odot}$ .

In table 3 are listed some of the data which are significant in the analysis of the large-scale physical constitution of the three clusters of nebulae in Coma, in Hydra and in Perseus.

The diameters for the three clusters given in table 3 refer of course only to those distances from the center of the clusters at which the emergence of the brighter cluster nebulae from the background nebulae can be statistically ascertained. When sufficient observations with more powerful telescopes are available so that the fainter member nebulae of the clusters can also be included in the counts it will probably be found that the diameters of the clusters are still larger than those given in table 3.

Unfortunately only a few more spherically symmetrical clusters can be reached with the 18-inch Schmidt telescope and these are considerably less rich in nebulae than those listed in table 3. The important question of the segregation of nebulae of different brightness can also only partly be solved until larger Schmidt telescopes are available. In addition, the velocity distribution as a function of the brightness of the nebulae must be investigated. The problems of the segregation of nebulae and of the dispersion in the velocities demand also further theoretical clarification since the classical statistical mechanics gives no answer to the fundamental questions which arise in connection with gravitational coöperative assemblies which are composed of such diverse elements as the nebulae, the stars and the constituents of intergalactic and interstellar matter. The uniformity of the structure of symmetrical clusters of nebulae observed so far and the quantitative agreement of their observed physical characteristics with those derived theoretically for the isothermal gravitational "gas" sphere suggest that the short-time scale associated customarily with the hypothesis of an expanding universe will perhaps become the less attractive the more the investigations on the large-scale distribution of matter in the universe progress.

<sup>1</sup> Zwicky, F., *Proc. Nat. Acad. Sci.*, **28**, 317-320 (1942).

<sup>2</sup> Zwicky, F., *Ibid.*, **28**, 150 (1942); *Astrophys. Jour.*, **95**, 555 (1942).

<sup>3</sup> Emden, R., *Gaskugeln*, Teubner, Leipzig, 1907.

<sup>4</sup> Hubble, E., *Astrophys. Jour.*, **71**, 231 (1930).

<sup>5</sup> Zwicky, F., *Th. von Kármán Anniversary Volume*, May, 1941, p. 137.

## THE EPIDEMIC CURVE

BY EDWIN B. WILSON AND MARY H. BURKE

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated July 15, 1942

In 1929 Soper<sup>1</sup> developed a theory of the epidemic curve based on tracing the rise and fall of the disease by "generations" of successive groups of infectious cases. If in the  $i$ th generation there are  $C_i$  infectious cases and  $S_i$  susceptibles and if there be a constant number  $A$  of susceptibles per infectious generation coming into the population, his equations to determine the number of cases and the number of susceptibles in the next generation were<sup>2</sup>

$$\frac{C_{i+1}}{C_i} = \frac{S_i}{m}, \quad (1a)$$

and

$$S_{i+1} = S_i - C_{i+1} + A, \quad (1b)$$

where  $m$  is the number of susceptibles necessary for one old case to produce just one new case. In the stationary endemic condition  $C_{i+1} = C_i$ ,  $S_i = m$ ,  $S_{i+1} = S_i$  and  $C_{i+1} = A$ . The length of the generation was taken to be the "incubation period." Thus if a fortnight be taken as the incubation period for measles and if there are 150 children per fortnight coming into the population through birth and growing up (with due allowance for deaths and emigration and immigration) measles could maintain itself in a stationary endemic condition with 150 cases per fortnight, the quantity  $m$  would have to be determined by enumerating the susceptibles in the population under such conditions, and might well be of the order of 4 to 5 years' worth of recruits  $A$ , which at 150 per fortnight would be 15,600 to 19,500. As measles does not occur in a steady endemic condition but in sharp epidemics, it is necessary according to Soper's equations that  $m$  should be variable or that there should be an accumulation of susceptibles to a number considerably greater than  $m$  before the epidemic, with an alternative deficit of susceptibles considerably below  $m$  after the epidemic.

Already in February, 1928, Dr. Wade H. Frost when delivering the Cutter Lectures in Preventive Medicine at the Harvard Medical School

had presented a similar theory based on a somewhat different line of thought.<sup>8</sup> He, too, considered that for the diseases to which his theory would be applicable one could think in terms of generations of infectious persons  $C_t$  who being in contact with susceptibles  $S_t$  for a certain period would infect some of them who would then become the new generation of infectious persons at an average time later by the "incubation period"; but he recognized that in the intermixture of susceptibles and infectious some of the susceptibles might have multiple contacts with infectious persons and yet could develop at most only one infection apiece. If  $S_t$  were the number of susceptibles,  $p = 1/S_t$  would be the chance that any particular contact would fall upon any particular susceptible and  $q = 1 - 1/S_t$  would be the chance that he would escape. If then there were  $k_t$  contacts made between infectious and susceptibles the chance that a susceptible would escape them all would be  $q^{k_t}$  and the chance that he would have at least one contact would be  $1 - q^{k_t}$ . The number of infected would therefore be  $S_t(1 - q^{k_t})$ . Frost assumed that the number of contacts  $k_t$  would be proportional to the numbers of infectious and of susceptibles jointly, or  $k_t = rC_tS_t$ . Thus his equations corresponding to Soper's 1a and 1b are

$$C_{t+1} = S_t \left[ 1 - \left( 1 - \frac{1}{S_t} \right)^{rC_tS_t} \right], \quad (2a)$$

and

$$S_{t+1} = S_t - C_{t+1} + A, \quad (2b)$$

except that Frost did not allow for the recruitment of susceptibles at the rate  $A$  per incubation period for the reason that he was satisfied at the time to give a theory of the curve of an epidemic so sharp that the number of recruits during the epidemic would not materially influence the course of the epidemic.

It is clear that any such theory as that proposed by Soper or by Frost cannot be expected to explain in quantitative detail the course of any epidemic; any precise theoretical discussion of the epidemic curve must be highly mathematical and difficult and hypothetical. The greatest immediate value of the development of the theory and of attempts at its application to concrete instances must be upon the qualitative side in indicating the sorts of things which may happen under various idealized conditions. It is noteworthy that Soper's paper did accomplish instructive results and that Frost's discussion in his lectures here, but principally at the Johns Hopkins School of Hygiene and Public Health with his students in successive years, has likewise been deemed of great value. It is noteworthy also that although the two theories seem to be different as exemplified in equations (1) and (2) they are in fact pretty much alike. If the

number of infectious  $C_i$  and the contact rate  $r$  are small enough so that only the first two terms of the expansion of  $(1 - 1/S_i)^{rC_iS_i}$  need be taken, (2a) becomes

$$C_{i+1} = S_i[1 - (1 - rC_i)] = rC_iS_i$$

and  $r = 1/m$  makes (2a) then the same as (1a) even though Soper apparently did not think of  $1/m$  as a contact rate.<sup>4</sup>

As the formula (2a) is not easy to compute because of the high powers of numbers very near to 1 to which it leads, it was suggested to Dr. Frost that the "law of small numbers" could be used to modify the formulae without material change in the results so long as the number of susceptibles did not decline too far. Thus  $(1 - 1/S)^{rCS} = e^{-rC}$  and

$$C_{i+1} = S_i[1 - e^{-rC_i}], \tag{3a}$$

and

$$S_{i+1} = S_i - C_{i+1} + A. \tag{3b}$$

Table 1 gives the calculation of an epidemic where a single infectious case  $C_0 = 1$  is introduced into a population of  $S_0 = 2000$  susceptibles under the hypothesis that the rate of effective contact of infectious and susceptibles is  $r = 0.001$  (each infectious person averages 2 contacts with susceptibles), where the recruitment  $A$  is neglected and where the results of formulae (1), (2) and (3) are compared, keeping calculations to the nearest integer.

TABLE 1  
COURSE OF HYPOTHETICAL EPIDEMIC,  $C_0 = 1, S_0 = 2000, r = 0.001$

GENERATION	$m = 1000$ FORMULAE (1)	$r = 0.001$ FORMULAE (2)	$r = 0.001$ FORMULAE (3)
1	2	2	2
2	4	4	4
3	8	8	8
4	16	16	16
5	32	32	32
6	62	61	61
7	116	111	111
8	204	186	186
9	317	267	268
10	393	308	309
11	332	265	267
12	171	173	173
13	59	90	89
14	17	41	40
15	5	18	17
16	1	8	7
17		3	3
18		1	1
Total infected	1739	1594	1594
Residual susceptibles	261	406	406



It is clear that the results of calculations by (2) and (3) are essentially identical, and that the elimination of the double contacts makes the epidemic longer, more symmetrical and lower at the peak, and leaves more susceptibles untouched at the end.<sup>5</sup>

If the formulae (3) are used a very neat result may be had for the relation between  $S_0/m$ , the ratio of initial susceptibles to the number  $m = 1/r$ , and  $S_E/m$ , the ratio of the number of residual susceptibles to the same number. Indeed

$$S_1 = S_0 e^{-rC_0}, \quad S_2 = S_1 e^{-rC_1}, \quad \dots \quad S_{k+1} = S_k e^{-rC_k} \quad (4)$$

Multiplying these together and cancelling  $S_1, \dots, S_k$ ,

$$S_{k+1} = S_0 e^{-r(C_0 + C_1 + \dots + C_k)} \quad (5)$$

In any epidemic to which the theory applies the initial number of cases  $C_0$  introduced into the population of susceptibles would be few compared with the number around the peak, and the terminal is 0 after the epidemic has passed.<sup>6</sup> Hence if  $S_E$  be the number of susceptibles left, the total cases  $C_0 + C_1 + \dots + C_k$  must be essentially  $S_0 - S_E$ , and one has the result

$$S_E = S_0 e^{-r(S_0 - S_E)} \text{ or } \frac{rS_E}{rS_0} = e^{-rS_0(1 - S_E/S_0)} \quad (6)$$

where  $rS_0 = S_0/m$  and  $rS_E = S_E/m$ . If  $F = S_0/m$  and  $f = S_E/m$ ,

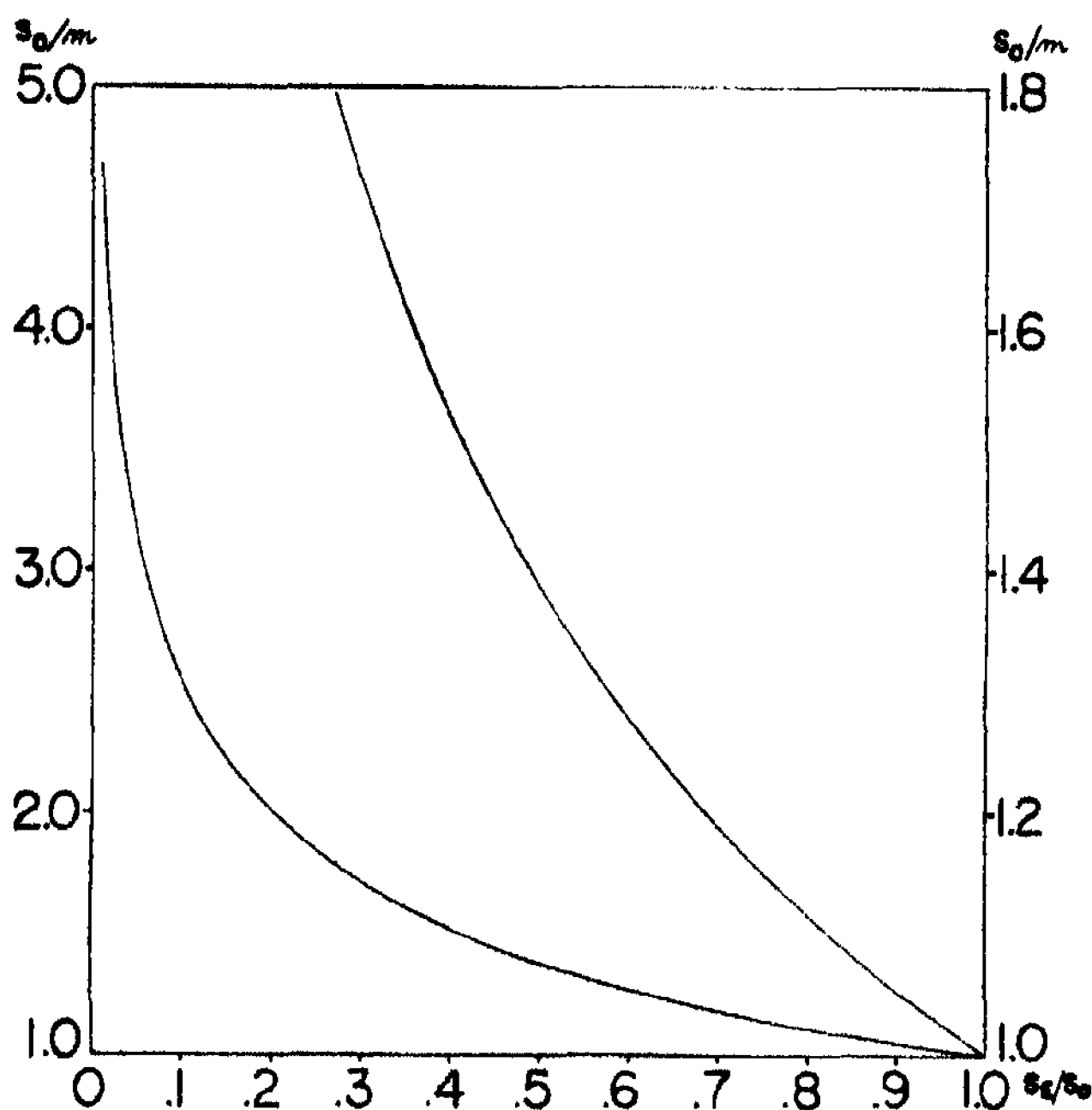
$$f/F = e^{-F(1 - f/F)} \quad (7)$$

This relation (7) between  $f$  and  $F$  cannot be solved for either, but a table of corresponding values of  $F$  and  $f$  may be computed and tabulated as in table 2. The results of the table are given in the figure.

TABLE 2  
RELATION BETWEEN THE RATIOS  $F$  AND  $f$

$\frac{f}{F} = \frac{S_E}{S_0}$	$F = \frac{S_0}{m}$	$f = \frac{S_E}{m}$	$\frac{f}{F} = \frac{S_E}{S_0}$	$F = \frac{S_0}{m}$	$f = \frac{S_E}{m}$
0.01	4.652	0.04652	0.32	1.676	0.5362
0.02	3.992	0.07984	0.35	1.615	0.5653
0.03	3.615	0.1084	0.40	1.527	0.6109
0.04	3.353	0.1341	0.45	1.452	0.6533
0.05	3.153	0.1577	0.50	1.386	0.6932
0.07	2.859	0.2002	0.55	1.329	0.7307
0.10	2.558	0.2558	0.60	1.277	0.7663
0.12	2.409	0.2891	0.65	1.231	0.8000
0.15	2.232	0.3348	0.70	1.189	0.8322
0.17	2.135	0.3629	0.75	1.151	0.8630
0.20	2.012	0.4024	0.80	1.116	0.8926
0.22	1.941	0.4271	0.85	1.083	0.9210
0.25	1.848	0.4621	0.90	1.054	0.9482
0.27	1.794	0.4843	0.95	1.026	0.9746
0.30	1.720	0.5160	1.00	1.000	1.0000

It is clear that if the number of susceptibles at the start were  $3.35 m$ , only 4% of the original susceptibles would remain untouched by the epidemic, 96% would contract the disease. In Panum's Faroe Islands epidemic of measles something like 96% of the population of the villages



Plot of relation between the fraction  $S_E/S_0$  of residual susceptibles to initial susceptibles and the multiple  $S_0/m$  which initial susceptibles are of  $m = 1/r$  (scale on the left), with enlargement for a part of the range (scale on the right).

entered by the disease was actually attacked by it.<sup>7</sup> It is further clear from the table that if  $S_0/m$  were only about 2, the fraction of the susceptibles which would escape would be 20%. For Hedrich's analysis of measles epidemics in Baltimore,<sup>8</sup> the table would not be strictly applicable because there was recruitment of the population which for a fairly long-drawn out epidemic of some 8 months might not be negligible. For his epidemic of 1930-1931, the susceptibles at the beginning were 78,968 but they rose to 81,449 during the initial stages where recruitment exceeded cases; from then they fell to 52,111 before the end of the epidemic, rising to 54,408 at its end. The ratio of minimum to maximum is 0.64 whereas that of end to beginning is 0.69. If we enter the table with the value  $S_E/S_0 = 0.64$  we find  $S_0/m = 1.240$  which would give  $m = 65,300$  on the base  $S_0 = 81,000$ ;

if we enter with  $S_E/S_0 = 0.69$  we find  $S_0/m = 1.199$  which would give  $m = 66,000$  on the basis of  $S_0 = 79,000$ . These estimates of  $m$  are nearly alike. There were about 13,000 children coming into the susceptible group each year, which means that the number of susceptibles  $m = 1/r$  would correspond to about 5 years of births, which may not be an unreasonable result in view of the certainty that both the theory and the calculations can at best be regarded as only approximately representing real conditions.

<sup>1</sup> Soper, H. E., "The Interpretation of Periodicity in Disease Prevalence," *J. Roy. Statist. Soc. London*, 92, 34-61 (1929).

<sup>2</sup> We shall not restate in detail the conditions under which the theory might be considered as approximately true, nor at this time enter upon a discussion of the question of periodicity. The rigorous equations for a theory involving the basic conceptions and restrictions would seem to be these: First, there are at any time a number of infectious persons  $I(t)$  and a number of susceptibles  $S(t)$ . Second, the rate of loss of susceptibles is  $-dS/dt$  and should be set equal to the rate at which susceptibles become infected, namely,  $C(t)$ , less the rate of recruitment of new susceptibles  $A(t)$ . Third, the rate  $C(t)$  is taken to be proportional to the product of  $I(t)$  and  $S(t)$ . Fourth, the newly infected persons  $C(t)dt$  become infectious after a latent time  $\tau$  and remain infectious for a time  $\sigma$ . Hence

$$-\frac{dS}{dt} = C(t) - A(t), \quad C(t) = r(t)I(t)S(t), \quad I(t) = \int_{t-\tau-\sigma}^{t-\tau} C(t)dt. \quad (A)$$

The factor  $r$  and the rate of recruitment  $A$  are generally taken as constant, though Soper shows that probably  $r$  or his  $m$  which is the reciprocal of  $r$  has a seasonal variation. Here the symbols  $C$  and  $A$  are rates, instead of being numbers of individuals as in (1). The variables  $C$  and  $I$  may be eliminated to get

$$A - \frac{dS}{dt} = rS(t) \int_{t-\tau-\sigma}^{t-\tau} \left( A - \frac{dS}{dt} \right) = rS(t) [A\sigma - S(t-\tau) + S(t-\tau-\sigma)] \quad (B)$$

which is a differential-difference equation for  $S$ .

<sup>3</sup> Dr. Frost's lectures were on Feb. 2-3, 1928, and the dates of my letters to him were Feb. 9, 23 with replies from him dated Feb. 14, Mar. 21. It was in my second letter that I suggested the use of the law of small numbers in the way mentioned below. I strongly urged Dr. Frost to publish his theory of the epidemic curve, but he thought it too slight a contribution.—E. B. W.

<sup>4</sup> If we take Frost's equations and express the condition for a steady state we have  $C_t = C_{t+1} = A$  and  $A = S[1 - (1 - 1/S)^{rAS}]$ . This equation cannot be solved strictly for the relationship  $rS = 1$  with  $S = m$  for the steady state, the relation between  $r$ ,  $A$ ,  $S$  for the steady state is more complicated. If we make  $S$  constant in (B) we have  $1 = rS\sigma$  so that it is  $r\sigma$  which takes the place of  $r$  in Frost's theory or of  $1/m$  in Soper's; this is quite to be expected because the effective rate of generation of new cases must be the product of a contact rate  $r$  by a time  $\sigma$  available for making contacts. While for illustrative purposes to show what sorts of things may happen we may try different values of  $r$  or  $1/m$  in the Frost or Soper theories, and different values of  $C_0$  and  $S_0$ , and of  $A$  if we wish to admit recruitment, it must be remembered that in efforts to interpret concrete epidemics by the theory, the quantities  $r$ ,  $C_0$ ,  $S_0$ ,  $A$  have to be determined from the data and cannot be expected to be determinable except

within rather wide limits. As a matter of fact the contact rate  $r$  must be highly variable within any community because contact within the home, within the school, and within the community at large must be at very different rates, so that at best  $r$  or  $m$  must be a sort of over-all community average of such very different rates.

<sup>6</sup> One of the limitations of Soper's set-up is that if  $m$  is sufficiently small, the calculation runs into the impossible situation where  $C_i + 1$  becomes greater than the remaining  $S_i$ ; for example, with  $C_0 = 1$ ,  $S_0 = 2000$ , and  $m = 500$ , the cases in successive generations are 4, 16, 63, 243, 814 (by which time  $S = 860$ ), and the next value of  $C$  by (1) comes out at around 1400. Frost's method involving elimination of double contacts seems not to suffer from this defect; calculating by the law of small numbers as in (3) we find with  $r = 1/500$  for this case the successive values of  $C$  as 4, 16, 62, 224, 611, 764, 250, 27, 2, leaving 40 susceptibles still untouched. The calculations have been carried to tenths and then rounded off at the end in tabulating cases in successive generations. If we go through the detailed calculation with the Frost formulae (2) we find the same integral values of  $C$  as with formulae (3). When calculating according to Soper's formula (1a) we may be doing him an injustice; after giving that formula he shifted over to the formula

$$\frac{C_i + 1/2}{C_i - 1/2} = \frac{S_i}{m}, \text{ i.e., } \frac{\text{no. cases next interval}}{\text{no. cases last interval}} = \frac{\text{no. susceptibles at present}}{m}$$

"since the change in  $S_i$  is usually small in the unit interval." This shift is advantageous for the analytical developments upon which he is entering and surely makes no change which would not be within the tolerances of approximations in the theory as applied in concrete cases. It is impossible to determine from his paper whether he based his numerical calculations upon the original form (1a) or upon this modified form.

<sup>4</sup> Under the conditions, the epidemic has to die out, as it would not have to if there were recruitment; if one would add the initial cases  $C_0$  which were introduced into the susceptible population  $S_0$  to  $S_0$  itself and use  $S_0 + C_0$  in place of  $S_0$  in (6) it would not be necessary to disregard the small number  $C_0$  as mentioned in the text.

<sup>7</sup> Panum, P. L., *Observations Made During the Epidemic of Measles on the Faroe Islands in the Year 1846*, Delta Omega Society, 1940, distributed by the American Public Health Association, New York, N. Y.

<sup>8</sup> Hedrich, A. W., "Monthly Estimate of the Child Population Susceptible to Measles, 1900-1931, Baltimore, Md.," *Amer. J. Hygiene*, 17, 613-636 (1933).

*AUTOMORPHISMS OF THE DIHEDRAL GROUPS*

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated July 24, 1942

The group of inner automorphisms of a dihedral group whose order is twice an odd number is obviously the group itself while the group of inner automorphisms of a dihedral group whose order is divisible by 4 is the quotient group of this dihedral group with respect to its invariant subgroup of order 2 when the dihedral group is not the four group. It is well known that the four group is the only abelian dihedral group and that its group of automorphisms is the symmetric group of order 6. All of these automorphisms except the identity are outer automorphisms but this symmetric group admits no outer automorphisms. In fact, we shall prove that it is the only dihedral group which does not admit any outer automorphisms. To emphasize this fact it may be noted here that on page 152 of the *Survey of Modern Algebra* by Birkhoff and MacLane (1941) it is stated that the group of symmetries of the square, which is also a dihedral group, admits no outer automorphisms. This is obviously not in agreement with the theorem noted above.

To prove that the symmetric group of order 6, which is also the group of movements of the equilateral triangle, is the only dihedral group which does not admit any outer automorphisms it may first be noted that if a dihedral group whose order is twice an odd number admits no outer automorphisms its cyclic subgroup of index 2 cannot have an order which exceeds 3 since in the group of automorphisms of a cyclic group every operator of highest order corresponds to every other such operator and every operator may correspond to its inverse. If a dihedral group whose order is divisible by 4 admits no outer automorphisms it cannot be the four group and hence it contains a cyclic subgroup which involves an invariant subgroup of order 2. Its non-invariant operators correspond to themselves multiplied by every operator of this subgroup in some automorphism of the group. That is, *the dihedral group of order 6 is the only dihedral group which admits no outer automorphisms.*

Since the dihedral group of order 6 involves no invariant operator besides the identity and admits no outer automorphisms it is its own group of automorphisms. To prove that no other dihedral group whose order is twice an odd number is its own group of automorphisms it may be noted that if such a group has this property its cyclic subgroup of index 2 cannot admit any automorphisms besides the one in which each of its operators corresponds to its inverse and hence this cyclic subgroup of odd order must be of order 3 and hence the group must be the dihedral group of order 6.

If a dihedral group whose order is divisible by 4 is its own group of automorphisms its cyclic subgroup of index 2 may be assumed to have an order which exceeds 2 and to admit also no automorphism besides the identity and the one in which each of its operators corresponds to its inverse. Hence this cyclic group is of order 4 and it results that *the octic group and the symmetric group of order 6 are the only two dihedral groups which are their own groups of automorphisms.*

If a dihedral group whose order is twice an odd number admits the same number of outer automorphisms as inner automorphisms its cyclic subgroup of index 2 must have the cyclic group of order 4 for its group of automorphisms and hence it is the metacyclic group of order 20. This results from the fact that the group of automorphisms of the cyclic group of prime order  $p$  is the cyclic group of order  $p-1$ . If a dihedral group whose order is divisible by 4 admits the same number of outer automorphisms as inner automorphisms and is not the four group its cyclic subgroup of index 2 cannot have a larger order than 4 since its non-invariant operators of order 2 are transformed into themselves multiplied by the operators of the invariant subgroup of order 2 under its group of inner automorphisms, and hence they are transformed into themselves multiplied by the remaining operators of this cyclic subgroup by the outer automorphisms of the group. Hence it results that *there are two and only two dihedral groups which have the property that they admit just as many outer automorphisms as inner automorphisms.* One of these is the octic group while the other is the metacyclic group of order 20.

It was noted at the opening of this article that the group of inner automorphisms of a non-abelian dihedral group whose order is divisible by 4 is its quotient group with respect to its invariant subgroup of order 2. Hence every such dihedral group involves exactly two subgroups which are separately simply isomorphic with their groups of inner automorphisms while every other dihedral group except the four group is simply isomorphic with its group of inner automorphisms. On the other hand, the group of automorphisms of every non-abelian dihedral group is known<sup>1</sup> to be the holomorph of its cyclic subgroup of index 2. There is therefore no upper limit to the ratio of the orders of the group of automorphisms of the general dihedral group and the order of this dihedral group, and the study of the outer automorphisms of a dihedral group is practically reduced to the study of the holomorphs of cyclic groups.

Every abelian group  $H$  can be extended by operators which transform every operator of  $H$  into its inverse so as to obtain a group  $G$  of twice the order of  $H$  known as the generalized dihedral group of  $H$ . A necessary and sufficient condition that  $G$  is abelian is that all of its operators besides the identity are of order 2. As the group of inner automorphisms of  $G$  is its quotient group with respect to its subgroup generated by its invariant

operators it results that a necessary and sufficient condition that it is simply isomorphic with  $G$  is that  $H$  is of odd order. It therefore results that a necessary and sufficient condition that the group of inner automorphisms of the generalized dihedral group is simply isomorphic with itself is the same for both the dihedral group and the generalized dihedral group.

To prove the fact that it is not possible for a generalized dihedral group which is not also a dihedral group to be its own group of automorphisms it may first be noted that the order of such a generalized dihedral group could clearly not be twice an odd number since in that case it would be simply isomorphic with its group of inner automorphisms but would clearly also admit outer automorphisms since in every such inner automorphism an operator of odd order corresponds either to itself or to its inverse. Since every non-cyclic abelian group involves at least one non-cyclic Sylow subgroup and every non-cyclic Sylow group involves a non-identity automorphism in which not every operator corresponds to its inverse, it results that *there is no generalized dihedral group which is not also a dihedral group but which is its own group of automorphisms*. It was noted above that such a group may be its own group of inner automorphisms.

To prove that the group of automorphisms of every dihedral group whose order is twice an odd number as well as the group of automorphisms of every non-abelian generalized dihedral group whose order is twice an odd number is a complete group, it may first be noted that each of these groups of automorphisms is known to be the holomorph of its invariant abelian subgroup. Moreover, this holomorph involves no invariant operator besides the identity. If this group of automorphisms is transformed into itself by an operator which is not contained in it the co-set with respect to the given group of automorphisms which contains this transforming operator contains also an operator which is commutative with every operator of the given dihedral group or of the given generalized dihedral group, respectively.

The latter operator can therefore be so selected that its first power which appears in the given group of automorphisms is the identity and that it is commutative with every operator of this group since no two of the operators of this group transform the operators of the given invariant dihedral group in the same manner. It may be emphasized here that while a group may be its own group of automorphisms *there is no group whose order exceeds 2 which is its own holomorph*. This results from the fact that the holomorph of a group necessarily includes the group and if a group is non-abelian its holomorph contains a subgroup which is simply isomorphic with it and is composed of operators which are separately commutative with every operator of the given group. The operation of forming successive holomorphs beginning with a group whose order exceeds 2 therefore leads to an infinite system of groups of increasing orders.

Since the group of automorphisms of a dihedral group whose order is divisible by 4 contains an invariant operator of order 2 it cannot be a complete group. Moreover, it does not follow in this case that if an operator is commutative with every operator of a dihedral group it is also commutative with every operator of its group of automorphisms, since two distinct operators of this group do not necessarily transform the operators of the given invariant dihedral group in a different manner when the order of this dihedral group is divisible by 4. The automorphisms of the dihedral groups are unusually well adapted for the study of the general properties of the group of automorphisms of a given group in view of the properties noted above.

<sup>1</sup> Miller, Blichfeldt, Dickson, *Finite Groups*, p. 169 (1916).

## SOME ASYMPTOTIC RELATIONS

BY H. BATEMAN

CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated July 21, 1942

$$1. \text{ Let } (1 - x + tx)^m = \sum_{n=0}^{\infty} a_n t^n, (1 - t)^{-r-1} = \sum_{n=0}^{\infty} A_n^{(r)} t^n \quad |t| < 1 \\ A_n^{(r)} S_n^{(r)} = A_n^{(r)} a_0 + A_{n-1}^{(r)} a_1 + \dots + a_n,$$

then the equation

$$(1 - x + tx)^m (1 - t)^{-r-1} = \sum_{n=0}^{\infty} A_n^{(r)} F(-m, -r; -n-r; x) t^n \quad (1)$$

mentioned in a previous paper,<sup>1</sup> indicates that

$$S_n^{(r)} = F(-m, -r; -n-r; x).$$

Hence

$$\lim_{n \rightarrow \infty} S_n^{(r)} = 1.$$

This means that the Cesàro sum  $(C, r)$  of the series  $\sum_{n=0}^{\infty} a_n$  is 1. Equation

(1) still holds for  $x = 1$  if  $m$  is a positive integer and we write

$$F(-m, -r; -n-r; 1) = n! \Gamma(n+r-m+1) / (n-m)! \Gamma(n+r+1)$$



The corresponding equation

$$\begin{aligned} A_n^{(r)} F(a, -r; -n-r; 1) &= A_{n+a}^{(r)} & n+a \geq 0 \\ &= 0 & n+a < 0 \end{aligned}$$

holds even if  $r+1$  is a positive integer. The zero value when  $n+a < 0$  is indicated by Euler's relation

$$F(a, -r; -n-r; x) = (1-x)^{-n-a} F(-n-r-a, -n; -n-r; x).$$

2. The same relation in the form

$$\begin{aligned} F(n+1, n+1; x+n+\frac{3}{2}; \frac{1}{2}) &= \\ (1-\frac{1}{2})^{1/2} x^{-n-1/2} F(\frac{1}{2}+\frac{1}{2}x, \frac{1}{2}+\frac{1}{2}x; n+\frac{1}{2}x+\frac{3}{2}; \frac{1}{2}) \end{aligned}$$

may be used to obtain an asymptotic expression for the integral

$$t_n(x) = \int_0^\infty e^{-xa} \operatorname{sech} a \tanh^n a da$$

considered in a previous paper.<sup>2</sup> In fact if the foregoing relation is applied to the expression

$$t_n(x) = n!/(x+1)(x+3)\dots(x+2n+1) F(n+1, n+1; \frac{1}{2}x + n + \frac{3}{2}; \frac{1}{2}),$$

it is found that when  $n$  is a large positive integer

$t_n(x) \sim 2^{-1/2-1/2x} B(n+1, \frac{1}{2}+\frac{1}{2}x) \sim (2n)^{-1/2-1/2x} \Gamma(\frac{1}{2}+\frac{1}{2}x)$  where  $B(x, y)$  is the Beta function. The usual form of Laplace's method is inapplicable to  $t_n(x)$  because the function  $\tanh a$  has its maximum value at one end of the range of integration. The hypergeometric function under consideration is also not included in the types  $F(a, b \neq n; c; x)$ ,  $F(a \neq n, b \neq n; c; x)$ ,  $F(a, b \neq nc; \pm n; x)$ , which O. Perron<sup>3</sup> has studied for large positive values of  $n$ . These functions are represented by him with the aid of generating functions.

The more general integral of Stieltjes<sup>4</sup>

$$t_{m,n}(x) = \int_0^\infty e^{-xa} \operatorname{sech}^m a \tanh^n a da \quad m \geq 0, n \geq 0$$

satisfies the recurrence relations

$$(m+n)t_{m,n+1}(x) - nt_{m,n-1}(x) + xt_{m,n}(x) = 0$$

and can be obtained as a coefficient in the expansion of the integral

$$I = \int_0^\infty e^{-ax} (\operatorname{ch} a - z \operatorname{sh} a)^{-m} = \sum_{n=0}^\infty (m, n) t_{m,n}(x) z^n / n!$$

Now

$$I = \left( \frac{2}{1-z} \right)^m \frac{1}{x+m} F\left( \frac{1}{2}x + \frac{1}{2}m, m; \frac{1}{2}x + \frac{1}{2}m + 1; \frac{z+1}{z-1} \right) = \frac{1}{x+m} F\left( 1, m; \frac{1}{2}x + \frac{1}{2}m + 1; \frac{1}{2} + \frac{1}{2}z \right).$$

Hence

$$t_{m,n}(x) = n!/(x+m)(x+m+2)\dots(x+m+2n)F(n+1, m+n; \frac{1}{2}x + \frac{1}{2}m + n + 1; \frac{1}{2}) = 2^{1/2 m - 1/2 x - 1} B(\frac{1}{2}x + \frac{1}{2}m, n+1) F(\frac{1}{2}x + \frac{1}{2}m, \frac{1}{2}x + 1 - \frac{1}{2}m; \frac{1}{2}x + \frac{1}{2}m + n + 1; \frac{1}{2}).$$

When  $n$  is large

$$t_{m,n}(x) \sim 2^{1/2 m - 1/2 x - 1} B(\frac{1}{2}x + \frac{1}{2}m, n+1) \sim 2^{m-1} (2n)^{-1/2 m - 1/2 x} \Gamma(\frac{1}{2}x + \frac{1}{2}m).$$

3. There is an extension of equation (1) involving Appell's hypergeometric function of two variables of the first type<sup>b</sup>

$$(1-t)^{a+b-c} [1-(1-x)t]^{-a} [1-(1-y)t]^{-b} = \sum_{n=0}^{\infty} A_n^{(c-1)} t^n F_1(-n; a, b; c; x, y) \quad |t| < 1, |xt| < |1-t|, |yt| < |1-t|,$$

where

$$F_1(d; a, b; c; x, y) = \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} (d, p+q)(a, p)(b, q)x^p y^q / \{(c, p+q)(p!)(q!)\}.$$

The equation indicates that if

$$[1-(1-x)t]^{-a} [1-(1-y)t]^{-b} = \sum_{m=0}^{\infty} e_m t^m$$

$$[1+xt/(1-t)]^{-a} [1+yt/(1-t)]^{-b} = \sum_{m=0}^{\infty} c_m t^m$$

then, for the series  $\sum c_m$  the Cesàro sum  $(C, c-1)$  is found from

$$S_n^{(c-1)} = F_1(-n; a, b; c; x, y)$$

and for the series  $\sum e_m$  the Cesàro sum  $(C, c-a-b-1)$  is found from

$$S_n^{(c-a-b+1)} = [A_n^{(c-1)} / A_n^{(c-a-b-1)}] F_1(-n, a, b; c; x, y).$$

<sup>a</sup> Bateman, H., *Proc. Nat. Acad. Sci.*, 26, 491-496 (1940).

<sup>b</sup> Bateman, H., *Tohoku Math. Jour.*, 37, 23-38 (1933).

<sup>3</sup> Perron, O., *Heidelberg Akad. Sitzungsber.* (1916), 9 Abh. 24 pp.; (1917), 1 Abh., 69 pp

<sup>4</sup> Stieltjes, T. J., *Quart. Jour. Math.*, **24**, 370-382 (1890); *Oeuvres*, **2**, 378-391.

<sup>5</sup> Appell, P., and Kampé de Fériet, J., *Fonctions hypergéométriques*, Paris, 1926.

## AN ORTHOGONAL PROPERTY OF THE HYPERGEOMETRIC POLYNOMIAL

BY H. BATEMAN

CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated July 21, 1942

1. Mittag-Leffler's polynomial  $g_n(z)$  has the orthogonal property

$$\int_{-\infty}^{\infty} g_m(-ix)g_n(ix)dx/xsh(x) = \begin{cases} 0 & m \neq n \\ 2/n & m = n \end{cases} \quad \begin{matrix} m > 0 \\ n > 0. \end{matrix} \quad (1.1)$$

This is readily obtained by inverting the integral representation<sup>1</sup>

$$g_n(ix) = (1/\pi) \sin(\pi x) \int_{-\infty}^{\infty} e^{ixu} (\tanh^{1/2} u)^n du / sh u \quad n \geq 1 \quad (1.2)$$

by means of Fourier's inversion formula. The resulting equation

$$\operatorname{cosech} u (\tanh^{1/2} u)^n = 1/2 \int_{-\infty}^{\infty} e^{-iux} \operatorname{cosech}(\pi x) g_n(ix) dx \quad (1.3)$$

then gives the desired relation when  $sh u e^{-iux}$  is expanded in powers of  $\tanh^{1/2} u$ . With the notation of the hypergeometric function the orthogonal relation may be written in the form

$$\int_{-\infty}^{\infty} F(1-m, 1+ix; 2; 2) F(1-n, 1-ix; 2; 2) \times \\ x dx / sh(\pi x) = \begin{cases} 0 & m \neq n \\ 1/2n, & m = n > 0. \end{cases} \quad (1.4)$$

2. A more general relation may be obtained by writing Euler's integral in the form

$$F(-n, ix; c; z) B(ix, c-ix) = \int_{-\infty}^{\infty} e^{ixu} z^{-1/2cu} (1 - 1/2z - 1/2z \tanh^{1/2} u)^n \\ du / (2ch^{1/2} u)^c \quad (2.1)$$

and treating it in much the same way as the integral used to represent  $g_n(ix)$ . The result is that if  $f_n(x) = F(-n, ix; c; z)$  where  $z$  is real

$$\int_{-\infty}^{\infty} B(ix, c-ix) (z-1)^{ix} f_m(x) f_n(x) dx = \begin{cases} 0 & m \neq n \\ (-)^n n! / (c, n) (z-1)^c + {}^n z^{-c} & m = n \end{cases} \quad (2.2)$$

where  $B(p, q)$  is the Beta function. This includes the relation

$$\begin{aligned} \int_{-\infty}^{\infty} \operatorname{sech} (1/2\pi x) w_m(x) w_n(x) dx &= 0 & m \neq n \\ &= 1 & m = n \end{aligned} \quad (2.3)$$

satisfied by the polynomial  $w_n(x)$  defined by means of the generating function

$$(1 + t^2)^{-1/2} \exp(-x \arctan t) = \sum_{n=0}^{\infty} w_n(x) t^n. \quad (2.4)$$

A polynomial with an orthogonal property for the weight function  $\operatorname{sech} (\pi x)$  was discovered in 1940 by G. H. Hardy.<sup>2</sup> My attention to the polynomial  $w_n(x)$  was called by a letter from B. R. Wicker of Loyola University, Los Angeles. He defined the polynomial in the first place by means of a definite integral equivalent to

$$i^n w_n(x) = (1/\pi) \int_{-\infty}^{\infty} e^{-ixu} \operatorname{sech} u \tanh^n u du \quad (2.5)$$

and by a contour integral. From these he obtained the generating function and a recurrence relation

$$(n+1)w_{n+1}(x) + nw_{n-1}(x) + xw_n(x) = 0. \quad (2.6)$$

He also noted the existence of an orthogonal relation but did not see the relation between his polynomial and that of Hardy. Originally Hardy used the notation  $q_n(x)$  and Wicker the notation  $Q_n(x)$  but as both of these notations are used for Legendre functions of the second kind the notation  $w_n(x)$  is preferable.

3. The polynomial  $w_n(x)$  may be expressed in terms of my polynomial<sup>3</sup>  $F_n(z)$  by means of the equation

$$i^n w_n(x) = \sum_{m=0}^n a_{nm} F_m(ix)$$

where

$$l^n = \sum_{m=0}^n a_{n,m} P_m(l). \quad (3.1)$$

If  $F_n^m(z)$  denotes Pasternack's polynomial<sup>4</sup> which is such that when  $R(m) > -1$  and  $|R(z)| < 1 + R(m)$

$$\begin{aligned} 2^m B(1/2m + 1/2 + 1/2iz, 1/2m + 1/2 - 1/2iz) F_n^m(iz) = \\ \int_{-\infty}^{\infty} e^{-ixz} P_n(\tanh x) \frac{dx}{\cosh^{m+1} x} \end{aligned} \quad (3.2)$$

we may deduce the orthogonal relation

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{F_n^m(iz) F_{n'}^{-m}(-iz) dz}{\operatorname{ch}(\pi z) + \cos(m\pi)} &= 0 & n' \neq n \\ &= 1/(2n+1) & n' = n. \end{aligned} \quad (3.3)$$

When  $m = 0$  we have  $F_n^m(z) = F_n(z)$  and the foregoing relation gives the known orthogonal property of  $F_n(z)$ . When  $m = 1$  the function  $F_n^m(z)$  reduces to the function  $E_n(z)$  for which an orthogonal relation has not yet been found. This polynomial  $E_n(z)$  was defined by the operational equation

$$E_n(d/du) \operatorname{sech}^2 u = \operatorname{sech}^2 u P_n(\tanh u). \quad (3.4)$$

When  $m = -1/2$  the orthogonal relation satisfied by  $F_n(iz)$  is of the same type as that satisfied by  $w_n(z)$  and, indeed, we have the relation

$$F_n^{-1/2}(1/2ix) = i^n w_n(x). \quad (3.5)$$

4. In the case of Rice's polynomial<sup>6</sup>  $H_n(x, p, v)$  which is represented by the integral

$$H_n(x, p, v) B(x, p-x) \doteq \int_{-\infty}^{\infty} e^{xt} (1+e^t)^{-p} P_n \left[ 1 - \frac{2ve^t}{1+e^t} \right] dt \quad (4.1)$$

the orthogonal relation seems to be

$$\begin{aligned} \int_{-\infty}^{\infty} H_n(iz, p, v) B(iz, p-iz) A_m(-iz, p, v) dz &= 0 & m \neq n \\ &= 2\pi & m = n \end{aligned} \quad (4.2)$$

where  $A_n(x, p, v)$  is defined by means of the expansion

$$\left( \frac{1-u}{u+2v-1} \right)^x \left( \frac{2v}{u+2v-1} \right)^p = \sum_{n=0}^{\infty} A_n(x, p, v) P_n(u). \quad (4.3)$$

It is readily found that

$$\begin{aligned} A_n(x, p, v) = (2n+1) &\left[ \frac{1}{x+1} F(x+p, x+1; x+2; v^{-1}) - \right. \\ &\{n(n+1)/1!1!\} \frac{1}{x+2} F(x+p, x+2; x+3; v^{-1}) + \\ &\{(n-1)n(n+1)(n+2)/2!2!\} \frac{1}{x+3} F(x+p, x+3; x+4; \\ &\left. v^{-1}) + \dots \right]. \end{aligned} \quad (4.4)$$

5. Hardy<sup>6</sup> has shown that the functions  $W_n(x) = \frac{\sin \pi(x-n)}{(x-n)}$  form an orthogonal system for the range  $-\infty : \infty$ , and it is natural to ask whether the functions  $b(x) = \operatorname{sech} x F_n\left(\frac{2ix}{\pi}\right)i^{-n}$  can be expressed as linear combinations of the functions  $W(x)$  by means of the formula of interpolation

$$b_m(x) = \sum_{n=-\infty}^{\infty} W_n(x)b_m(n). \quad (5.1)$$

This formula has been tested numerically for the case  $n = 0$ ,  $x = 1/2$  when it becomes

$$\pi \cdot \operatorname{sech} x = \sin \pi x \left[ \frac{1}{x} - \frac{2x}{x^2-1} \operatorname{sech} 1 + \frac{2x}{x^2-4} \operatorname{sech} 2 - \frac{2x}{x^2-9} \operatorname{sech} 3 + \frac{2x}{x^2-16} \operatorname{sech} 4 - \dots \right] \quad (5.2)$$

Using the values  $\operatorname{sech} 1 = .64805 \ 42734 \ 04663 \ 6$   
 $\operatorname{sech} 2 = .27111 \ 82739 \ 42365 \ 7$   
 $\operatorname{sech} 3 = .09132 \ 79274 \ 19433 \ 4$   
 $\operatorname{sech} 4 = .03661 \ 89934 \ 73686 \ 5$   
 $\operatorname{sech} 5 = .01334 \ 05293 \ 99091 \ 5$

$$\operatorname{ch}^{1/2} = 1.12762 \ 59652 \ 06381, \ \pi = 3.14159 \ 26535 \ 89793$$

derived from numbers in the British Association Tables, vol. 1, we find that  $\pi \operatorname{sech} 1/2 = 2.786 \dots$  while the right-hand side exceeds 2.791774 when only the first three terms are taken into consideration. The complete value is greater than this and so this numerical test indicates that the supposition (5.1) is false.

<sup>1</sup> Bateman, H., *Proc. Nat. Acad. Sci.*, **26**, 491-496 (1940).

<sup>2</sup> Hardy, G. H., *Proc. Cambridge Phil. Soc.*, **36**, 1-8 (1940).

<sup>3</sup> Bateman, H., *Tôhoku Math. Jour.*, **37**, 23-38 (1933); *Annals of Math.*, **35**, 767-775 (1934).

<sup>4</sup> Pasternack, S., *Phil. Mag.*, (7) **28**, 209-226 (1939).

<sup>5</sup> Rice, S., *Duke Math. Jour.*, **6**, 108-119 (1940).

<sup>6</sup> Hardy, G. H., *Proc. Cambridge Phil. Soc.*, **37**, 331-348 (1941).

## CONTINGENCY TABLES

BY EDWIN B. WILSON AND JANE WORCESTER

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated June 27, 1942

If there be given a fourfold universe such as exists in the case of a population having two characters  $A$  and  $B$ , with four probabilities  $p_1 = p_{AB}$ ,  $p_2 = p_{\alpha B}$ ,  $p_3 = p_{A\beta}$ ,  $p_4 = p_{\alpha\beta}$  whose sum is 1 for the presence or absence of the characters in pairs, and if a sample of  $N$  be drawn from it and tabulated in a fourfold table with  $n_1 = (AB)$ ,  $n_2 = (\alpha B)$ ,  $n_3 = (A\beta)$ ,  $n_4 = (\alpha\beta)$  in the notation of Yule<sup>1</sup> with  $\Sigma n = N$ , the chance of any particular table is

$$P = \frac{N! p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}}{n_1! n_2! n_3! n_4!} \quad (1)$$

By virtue of the definitions of probability, the probabilities of  $A$  and of  $B$  are, respectively,

$$p_A = p_1 + p_3, \quad p_B = p_1 + p_2. \quad (2)$$

There are many such tables, in fact  $(N+1)(N+2)(N+3)/6$ . If we observe a specific table in which

$$(A) = n_1 + n_3, \quad (B) = n_1 + n_2, \quad N = n_1 + n_2 + n_3 + n_4 \quad (3)$$

and if we limit our consideration of the  $(N+1)(N+2)(N+3)/6$  tables to only that small series of tables for which (3) holds between  $n_1$ ,  $n_2$ ,  $n_3$ ,  $n_4$ , then only one of the four numbers  $n_i$  is free to vary and the number of tables in the series is only one more than the least of the quantities  $(A)$ ,  $(B)$ ,  $(\alpha)$ ,  $(\beta)$ .

Leaving  $n_1$  undetermined and eliminating  $n_2$ ,  $n_3$ ,  $n_4$  from (1) by (3) we may write

$$P = \frac{N! p_1^{n_1} p_2^{(B) - n_1} p_3^{(A) - n_1} p_4^{N - (A) - (B) + n_1}}{n_1! n_2! n_3! n_4!} \quad (4)$$

The probabilities  $p_1$ ,  $p_2$ ,  $p_3$ ,  $p_4$  in the universe are unknown save that their sum must be 1. The value of  $P$  may, however, be written as

$$P = \frac{N! p_2^{(B)} p_3^{(A)} p_4^{N - (A) - (B)}}{n_1! n_2! n_3! n_4!} \left[ \frac{p_1 p_4}{p_2 p_3} \right]^{n_1} \quad (5)$$

As the first fraction of the product contains in the exponents of  $p_2$ ,  $p_3$ ,  $p_4$  only quantities which are taken as given by (3) for all the tables which are being considered, viz., all with fixed marginal totals, it will not vary from table to table in this series by virtue of the values  $p_2$ ,  $p_3$ ,  $p_4$ , no matter

what those unknown values be, but only by virtue of the changes in the value of  $n_1$  and the correlative changes of  $n_2, n_3, n_4$  which occur as factorials in the denominator of this fraction. The second factor, however, contains  $n_1$  which changes and will be invariant for the series of tables if and only if

$$p_1 p_4 = p_2 p_3. \quad (6)$$

In view of the identical conditions (2) taken with  $\Sigma p = 1$ , we may eliminate  $p_2, p_3, p_4$  from (6) leaving an equation in  $p_1$  alone, viz.,

$$p_1(1 + p_1 - p_A - p_B) = (p_B - p_1)(p_A - p_1) \quad (7)$$

or

$$p_1 = p_A p_B. \quad (8)$$

Thus if we specify that (5) be invariant in so far as it depends upon  $p_i$  from term to term in the series of tables for which marginal totals are given, we arrive at the condition that  $p_{AB} = p_A p_B$ , namely, that  $A$  and  $B$  be independent characters in the universe.

If we had started with the assumption that  $p_1 = p_A p_B$  then

$$P = \frac{N! p_A^{(A)} p_B^{(B)} p_a^{(\alpha)} p_b^{(\beta)}}{n_1! n_2! n_3! n_4!} \quad (9)$$

and the value of  $P$  within the series would depend in a constant manner on  $p_A$  and  $p_B$ . That which the proof given in formulas (1) to (8) shows is that conversely the relation  $p_1 = p_A p_B$  follows from the general assumption that  $P$  must not vary from term to term in the series because of the unknown probabilities but only because of the variation in  $n_1$ .

If our concern with respect to inference from the observations to the conclusion that they do or do not come from an associated universe may be limited to a calculation of relative frequencies in this series, we find that those relative frequencies are inversely as the product of the factorials of the numbers in the table; if we are willing to compute the products of the reciprocals of the factorials for all terms of the series, the relative probabilities may be found, but if the products of the reciprocals of the factorials may be summed algebraically, then the formula

$$R = \left[ \sum \frac{1}{n_1! n_2! n_3! n_4!} \right]^{-1} \frac{1}{n_1! n_2! n_3! n_4!} \quad (10)$$

may be used to compute the relative probabilities of the limited number of terms in the tail or tails of the series necessary to determine the significance or non-significance of the observations in the matter of association of  $A$  and  $B$  in their universe.



If we had started with samples of  $n$  and  $N - n$  in two different point-binomial universes with probabilities  $p_1 = p$ ,  $p_2 = 1 - p$ ,  $p_3 = p'$ ,  $p_4 = 1 - p'$  in the two pairs of cells coming from the two universes independently,<sup>2</sup> the probability of the observations would be

$$P = \frac{n! p_1^{n_1} p_2^{n_2}}{n_1! n_2!} \cdot \frac{(N - n)! p_3^{n_3} p_4^{n_4}}{n_3! n_4!} \quad (1')$$

The definitions of probability make

$$p_1 + p_2 = 1, \quad p_3 + p_4 = 1, \quad (2'a)$$

and furthermore if we add conditions (3) fixing the marginal totals  $n_1 + n_3$  and  $n_2 + n_4$ , of which the latter is not independent of the former,

$$(A) = n_1 + n_3, \quad (3')$$

and finally the probability of an  $A$  in sampling from the two independent universes must be

$$P_A = \frac{n}{N} p_1 + \frac{N - n}{N} p_3 \quad (2'b)$$

which has to be adjoined to (2'a) so that both (2'a) and (2'b) taken together correspond to (2). From these equations we may get

$$P = \frac{n! (N - n)! p_1^{n_1} p_2^{n_2 - n_1} p_3^{(A) - n_1} p_4^{N - (A) - n_2 + n_1}}{n_1! n_2! n_3! n_4!} \quad (4')$$

from which follows

$$P = \frac{n! (N - n)! p_1^{n_1} p_3^{(A)} p_4^{N - (A) - n_1} \left[ \frac{p_1 p_4}{p_2 p_3} \right]^{n_1}}{n_1! n_2! n_3! n_4!} \quad (5')$$

Introducing now the assumption that (5') may not vary from term to term in the short series selected from the  $(n + 1)(N - n + 1)$  possible samples by the restriction that the marginal totals  $(A)$  and  $N - (A)$  shall be as observed, we find again

$$p_1 p_4 = p_2 p_3 \quad (6')$$

which by virtue of (2'a) becomes

$$p_1(1 - p_3) = (1 - p_1)p_3 \quad \text{or} \quad p_1 = p_3 \quad (8')$$

and shows that the proportions in the two point-binomial universes must be the same; it has not been necessary to use (2'b) but this relation fixes  $p_1$  and  $p_3$  as equal to  $p_A$ . The relative probabilities within the series again reduce therefore to (10) in which the summation factor must be the same because the summation must be over the same values of  $n_i$  as before.<sup>3</sup>

The general principle seems to be capable of statement as:

If there be given  $v$  cellular universes,  $U_1, U_2, \dots, U_v$ , with unknown probabilities  $p_{11}, \dots, p_{1k_1}; p_{21}, \dots, p_{2k_2}; \dots; p_{v1}, \dots, p_{vk_v}$  to the total number of  $T$  in their cells; and if samples of  $N_1, N_2, \dots, N_v$  be drawn from them with numbers  $n_{11}, \dots, n_{1k_1}; n_{21}, \dots, n_{2k_2}; \dots; n_{v1}, \dots, n_{vk_v}$ ; and if certain independent totals to the number of  $L$  among the  $n$ 's be assigned including the  $v$  totals  $N_i$ , the condition that the probability of the samples drawn depends in the same way upon the unknown probabilities throughout the subseries of tables for which these  $L$  totals remain fixed, and which therefore has  $T - L$  degrees of freedom, will enable the probability of the samples within the series to be written in terms of  $L$  unknown probabilities and the  $L$  assigned totals and will give  $T - L$  relations between the unknown probabilities. The relative probabilities within the subseries of  $T - L$  degrees of freedom are then proportional to the reciprocals of products of the factorials of the numbers occurring in the cells of the tables and may be determined by computing these values for all tables in the series whether or not an algebraic formula for the sum of those reciprocals can be obtained. The  $T$  unknown probabilities may be expressed in terms of  $L - v$  unknown independent variables, thus determining the type of universe which represents the appropriate null hypothesis. If the  $L - v$  unknown variables be assigned values derived from the  $L$  relations among the  $n$ 's, the unknown probabilities and "expected values" in the samples may be estimated, from which the value of  $\chi^2$  for the observed table and an approximate value of the probability of a table departing from the table of expected values as much or more than the one observed may be had.

Whether so general a principle is of much utility is difficult to say. It will allow problems to be solved when the probability set-up is not at first clear. For example: Suppose there be given a universe of association between two characters and that a sample of  $N$  is drawn subject to the condition that  $(A) + (B) = M$  shall be fixed, i.e., the total number of individuals with character  $A$  or  $B$  or both shall be constant; then

$$P = \frac{N! p_1^{n_1} p_2^{n_2} p_3^{n_3} p_4^{n_4}}{n_1! n_2! n_3! n_4!} = \frac{N! p_1^{n_1} p_2^{M - (A) - n_1} p_3^{(A) - n_1} p_4^{N - M + n_1}}{n_1! n_2! n_3! n_4!}$$

$$= \frac{N! p_2^M p_4^{N - M}}{n_1! n_2! n_3! n_4!} \left[ \frac{p_1 p_4}{p_2 p_3} \right]^{n_1} \left[ \frac{p_3}{p_2} \right]^{(A)}$$

Hence, by the principle,

$$p_1 p_4 = p_2 p_3, p_2 = p_3 \text{ or } p_1 p_4 = p_2^2 = p_3^2$$

$$p_1 + p_2 + p_3 + p_4 = 1, 2p_1 + p_2 + p_3 = p_M$$

Note that  $p_M$  which is  $M/N$  need not be a probability at all if for no other reason than that it may exceed 1. Finally,

$$p_1 = \frac{1}{4} p_M^2, p_2 = p_3 = \frac{1}{2} p_M (1 - \frac{1}{2} p_M), p_4 = (1 - \frac{1}{2} p_M)^2$$

Thus if we had given a table with  $N = 8$  and  $M = 12$  as

$A$		$A$											
<table style="border-collapse: collapse; margin: 0 auto;"> <tr><td style="border: 1px solid black; padding: 2px 10px;">6</td><td style="border: 1px solid black; padding: 2px 10px;">0</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">0</td><td style="border: 1px solid black; padding: 2px 10px;">2</td></tr> </table>	6	0	0	2	6	<table style="border-collapse: collapse; margin: 0 auto;"> <tr><td style="border: 1px solid black; padding: 2px 10px;">4</td><td style="border: 1px solid black; padding: 2px 10px;">4</td></tr> <tr><td style="border: 1px solid black; padding: 2px 10px;">0</td><td style="border: 1px solid black; padding: 2px 10px;">0</td></tr> </table>	4	4	0	0	8	0	with
6	0												
0	2												
4	4												
0	0												
6	2	8	4	4	8								

as "expected" from  $n_i = Np_i$ , there would be nine possible tables in the series for which  $N = 8$  and  $M = 12$ , viz., the two written and these seven

4	3	5	2	4	2	5	1	5	0	4	1	4	0
1	0	0	1	2	0	1	1	2	1	3	0	4	0

The probabilities for these 9 in the order written are as

$$\frac{2}{2880}, \frac{5}{2880}, \frac{20}{2880}, \frac{12}{2880}, \frac{30}{2880}, \frac{24}{2880}, \frac{12}{2880}, \frac{20}{2880}, \frac{5}{2880}$$

or

$$\frac{2}{130}, \frac{5}{130}, \frac{20}{130}, \frac{12}{130}, \frac{30}{130}, \frac{24}{130}, \frac{12}{130}, \frac{20}{130}, \frac{5}{130}$$

whose sum is 1. Now the table written first has  $P = 0.015$  whereas for it  $\chi^2 = 8.00$  which gives  $P = 0.019$ ; and the table written second has accumulated probability for itself and all tables no more probable of  $P = 0.092$  whereas for it  $\chi^2 = 6.22$  corresponding to  $P = 0.046$ .

For this problem it might be difficult to set up a random sampling procedure and it might be difficult to perform algebraically the sum of  $[n_1! n_2! n_3! n_4!]^{-1}$  subject to the conditions  $\sum n_i = N$  and  $2n_1 + n_2 + n_3 = M$  or  $n_4 - n_1 = (\alpha\beta) - (AB) = N - M$ , but the principle gives a solution for the relative probabilities and for the expected table and if one is willing to accept that solution and the further rule that the significance of a table in the spread of two degrees of freedom is to be determined by the sum of the probabilities of the table and all tables no more probable, the two rules taken together will give a test of significance. Whether it is an asset or a liability to have rules which will do such things under such general conditions only time can tell. No rules for inference which have thus far been proposed, such as the rule of equal distribution of ignorance, the rule of

inverse probabilities, the rule based on confidence intervals or regions, and possibly even the Yates-Fisher rule, have stood a long test of time and remained acceptable.

<sup>1</sup> *An Introduction to the Theory of Statistics*, by G. U. Yule (or new edition by G. U. Yule and M. G. Kendall), Chaps. I-V, London, Charles Griffin and Co.

<sup>2</sup> Yates, F., *Suppl. Jour. Roy. Statist. Soc. London*, 1, 217-235 (1934). The proof we give is the converse of that given by R. A. Fisher, *Statistical Methods for Research Workers*, Art. 21.02, in which he assumes  $p = p'$ .

<sup>3</sup> In a recent note (these PROCEEDINGS, 28, 94-100 (1942) footnote 7) it was pointed out in discussing a  $2 \times 3$  table that if the observations

$n_1$	$n_2$	$n_3$	$n_1 + n_2 + n_3 = n$
$n_4$	$n_5$	$n_6$	$n_4 + n_5 + n_6 = N - n$

$$n_1 + n_4 = (A) \quad n_2 + n_5 = (B) \quad n_3 + n_6 = N - (A) - (B)$$

were made, the values of the relative probabilities involved would not depend in final analysis, according to the general principle adopted, on whether the table had arisen (1) from a true sixfold universe with probabilities  $\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6$ , with  $\Sigma\pi = 1$  from which  $\frac{(N+5)!}{5!N!}$  tables of  $N$  elements could be drawn, or (2) from a pair of trinomial universes from which  $n$  and  $N - n$  elements, respectively, were drawn with the six probabilities connected by  $\pi_1 + \pi_2 + \pi_3 = 1, \pi_4 + \pi_5 + \pi_6 = 1$  of which  $\frac{(n+2)!(N-n+2)!}{(2!)^2 n! (N-n)!}$  could be found or (3) from a set of three point-binomial universes from which  $(A), (B), N - (A) - (B)$  elements, respectively, were drawn with the six probabilities connected by  $\pi_1 + \pi_4 = 1, \pi_2 + \pi_5 = 1, \pi_3 + \pi_6 = 1$  of which

$$\frac{[(A)+1]![(B)+1]![N-(A)-(B)+1]!}{(A)!(B)![N-(A)-(B)+1]!}$$

could be found.

It is instructive to work out the second case in detail. We have

$$P = \frac{n!(N-n)! \pi_1^{n_1} \pi_2^{n_2} \pi_3^{n_3} \pi_4^{n_4} \pi_5^{n_5} \pi_6^{n_6}}{n_1! n_2! n_3! n_4! n_5! n_6!}, \quad (1')$$

$$\pi_1 + \pi_2 + \pi_3 = 1, \pi_4 + \pi_5 + \pi_6 = 1, \quad (2'a)$$

$$\pi_A = \frac{n}{N} \pi_1 + \frac{N-n}{N} \pi_4, \pi_B = \frac{n}{N} \pi_2 + \frac{N-n}{N} \pi_5, \quad (2'b)$$

$$(A) = n_1 + n_4, (B) = n_2 + n_5. \quad (3'')$$

By virtue of (3'')

$$P = \frac{n!(N-n)! \pi_1^{n_1} \pi_2^{n_2} \pi_3^{n_3} \pi_4^{n_4} \pi_5^{n_5} \pi_6^{n_6}}{n_1! n_2! n_3! n_4! n_5! n_6!} \quad (4')$$

$$= \frac{n!(N-n)! \pi_1^{n_1} \pi_4^{n_4} \pi_2^{n_2} \pi_5^{n_5} \pi_3^{n_3} \pi_6^{n_6}}{n_1! n_2! n_3! n_4! n_5! n_6!} \left[ \frac{\pi_1 \pi_6}{\pi_3 \pi_4} \right]^{n_1} \left[ \frac{\pi_2 \pi_6}{\pi_3 \pi_5} \right]^{n_2} \quad (5'')$$

Hence

$$\frac{\pi_1\pi_6}{\pi_3\pi_4} = 1, \quad \frac{\pi_2\pi_6}{\pi_3\pi_5} = 1, \quad (6'')$$

by virtue of the principle that for variation within the series of two degrees of freedom depending on  $n_1$  and  $n_2$ ,  $P$  must not change no matter what be the unknown probabilities in the universe or universes. Using next the relations (2''a) one finds

$$\begin{aligned} \pi_1(1 - \pi_4 - \pi_5) &= (1 - \pi_1 - \pi_2)\pi_4, & \pi_2(1 - \pi_4 - \pi_5) &= (1 - \pi_1 - \pi_2)\pi_5 \\ \pi_1(1 - \pi_5) &= (1 - \pi_2)\pi_4, & \pi_2(1 - \pi_4) &= (1 - \pi_1)\pi_5. \end{aligned}$$

Hence

$$\pi_1 = \pi_4, \pi_2 = \pi_5 \text{ and } \pi_3 = \pi_6,$$

and from this by (2''b) the probabilities  $\pi_1, \pi_2$  may be shown to be  $\pi_A, \pi_B$ . Then

$$P = \frac{n!(N-n)!(1-\pi_A-\pi_B)^n \pi_A^{(A)} \pi_B^{(B)} (1-\pi_A-\pi_B)^{N-(A)-(B)-n}}{n_1!n_2!n_3!n_4!n_5!n_6!}$$

which is precisely the expression that would have been written down in the first place if it had been assumed at the start that  $\pi_1 = \pi_4 = \pi_A$  and  $\pi_2 = \pi_5 = \pi_B$ .

The expression may be summed by writing down the probability of dividing  $N$  into  $(A), (B), N - (A) - (B)$ , viz.,

$$\frac{N! \pi_A^{(A)} \pi_B^{(B)} (1 - \pi_A - \pi_B)^{N-(A)-(B)}}{(A)!(B)![N-(A)-(B)]!}$$

Thus the relative probability within the series of two degrees of freedom is the quotient

$$R = \frac{(A)!(B)![N-(A)-(B)]!n!(N-n)!}{N!n_1!n_2!n_3!n_4!n_5!n_6!}$$

and would in fact be the same on either of the other two suppositions as to the origin of the observed members  $n_i$ .

## THE ASSOCIATION OF THREE ATTRIBUTES

BY EDWIN B. WILSON AND JANE WORCESTER

HARVARD SCHOOL OF PUBLIC HEALTH

Communicated August 14, 1942

If from a universe of individuals with or without three characters  $A, B, C$  there be drawn a sample of  $N$ , the six numbers  $(A), (B), (C), (\alpha), (\beta), (\gamma)$ , with or without each character will be known, as will the twelve numbers  $(AB), \dots, (\beta\gamma)$  of those with or without each pair of characters, and the eight numbers  $(ABC), \dots, (\alpha\beta\gamma)$  which specify the primary populations in the sample. Corresponding to these eight types of individuals there will be eight (presumably unknown) probabilities in the universe,  $p_{ABC}$ , etc.

The chance of drawing the particular sample is

$$P = \frac{N! p_{ABC}^{(ABC)} \dots p_{\alpha\beta\gamma}^{(\alpha\beta\gamma)}}{(ABC)! \dots (\alpha\beta\gamma)!} \quad (1)$$

When discussing the statistical significance of the sample, which may be considered as a  $2 \times 2 \times 2$  table, certain values taken from the sample must be "used." There are four cases which we propose to discuss:

1.  $N, (A), (B), (C)$ —the grand total and the "edge" subtotals.
2.  $(AB), (\alpha B), (A\beta), (\alpha\beta), (C), (\gamma)$ —one marginal "face" and the complementary edge.
3. Two marginal faces—those of  $A, B$  and  $A, C$ .
4. All three marginal faces.

In the discussion it will be assumed that the sample probability (1) must be independent of the values of the unknown cell probabilities for all variations of the sample consistent with the marginal totals used.<sup>1</sup>

1. *The Edge Subtotals and the Grand Total Used.*—There are four cell frequencies which may be assigned values consistent with these totals; they may be taken as<sup>2</sup>  $(ABC), (BC), (AC), (AB)$ . Then (1) becomes

$$\frac{N! p_{ABC}^{ABC} p_{\alpha BC}^{BC} \dots p_{AB\gamma}^{AB} \dots p_{A\beta\gamma}^{AC} \dots p_{\alpha\beta\gamma}^{N-A-B-C+AB+AC+BC-ABC}}{ABC! \alpha BC! A\beta C! AB\gamma! A\beta\gamma! \alpha B\gamma! \alpha\beta C! \alpha\beta\gamma!},$$

where the parentheses which Yule uses to designate numbers have been dropped, and will be reintroduced only when clarity requires it. The expression in  $p$  may be written as

$$\left( \frac{p_{ABC} p_{A\beta\gamma} p_{\alpha B\gamma} p_{\alpha\beta C}}{p_{AB\gamma} p_{A\beta C} p_{\alpha BC} p_{\alpha\beta\gamma}} \right)^{ABC} \left( \frac{p_{AB\gamma} p_{\alpha\beta\gamma}}{p_{A\beta\gamma} p_{\alpha B\gamma}} \right)^{AB} \left( \frac{p_{A\beta C} p_{\alpha\beta\gamma}}{p_{A\beta\gamma} p_{\alpha\beta C}} \right)^{AC} \times$$

$$\left( \frac{p_{\alpha BC} p_{\alpha\beta\gamma}}{p_{\alpha B\gamma} p_{\alpha\beta C}} \right)^{BC} \left( \frac{p_{A\beta\gamma}}{p_{\alpha\beta\gamma}} \right)^A \left( \frac{p_{\alpha B\gamma}}{p_{\alpha\beta\gamma}} \right)^B \left( \frac{p_{\alpha\beta C}}{p_{\alpha\beta\gamma}} \right)^C p_{\alpha\beta\gamma}^N$$

and if the probability is not to vary with  $ABC, AB, AC, BC$  one has

$$\frac{p_{AB\gamma} p_{\alpha\beta\gamma}}{p_{A\beta\gamma} p_{\alpha B\gamma}} = \frac{p_{A\beta C} p_{\alpha\beta\gamma}}{p_{A\beta\gamma} p_{\alpha\beta C}} = \frac{p_{\alpha BC} p_{\alpha\beta\gamma}}{p_{\alpha B\gamma} p_{\alpha\beta C}} = \frac{p_{ABC} p_{\alpha\beta\gamma}}{p_{A\beta C} p_{\alpha BC}} = 1.$$

These conditions taken with  $\Sigma p = 1$  express the interrelations of the eight  $p$ 's which are implied by the marginal totals used and the principle of independence assumed. The most suggestive way to solve in terms of three independent values of  $p$  seems to be to set

$$\frac{p_{AB\gamma}}{p_{\alpha\beta\gamma}} = \frac{p_1}{q_1}, \quad \frac{p_{\alpha B\gamma}}{p_{\alpha\beta\gamma}} = \frac{p_2}{q_2}, \quad \frac{p_{\alpha\beta C}}{p_{\alpha\beta\gamma}} = \frac{p_3}{q_3}$$

where  $p_i + q_i = 1$ . Then the first three conditions will give  $p_{AB\gamma}/p_{\alpha\beta\gamma}$ ,  $p_{A\beta C}/p_{\alpha\beta\gamma}$ ,  $p_{\alpha BC}/p_{\alpha\beta\gamma}$ , and the fourth  $p_{ABC}/p_{\alpha\beta\gamma}$ , whereupon  $\Sigma p = 1$  makes  $p_{\alpha\beta\gamma} = q_1 q_2 q_3$  and each probability becomes the product of three  $p$ 's and  $q$ 's.

The values of the cell probabilities remain unknown but expressed in terms of a reduced number, namely three, as  $p_{ABC} = p_1 p_2 p_3, \dots, p_{\alpha\beta\gamma} = q_1 q_2 q_3$ . There are four degrees of freedom. One may compute the relative probabilities of all the samples which may occur subject to the marginal totals specified; this may be done by writing down all the tables and taking their probabilities as inversely proportional to the product of the factorials of the cell frequencies, or it may be done by noting that the probability of the given totals  $A, B, C$  arising in three independent partitions of  $N$  is

$$\frac{N! p_1^\alpha q_1^\alpha N! p_2^\beta q_2^\beta N! p_3^\gamma q_3^\gamma}{A! \alpha! B! \beta! C! \gamma!} \quad (2)$$

and hence the relative probability of any table must be the quotient of (1) by (2) which reduces to<sup>3</sup>

$$\frac{A! \alpha! B! \beta! C! \gamma!}{(N!)^2 ABC! \dots \alpha\beta\gamma!} \quad (3)$$

If finally it is assumed that those tables which with all equally or less probable tables have a total (relative) probability no greater than 0.05 or some other assigned value are significant at the level of that value, the question of significance can be settled. Obviously from the form  $p_{ABC} = p_1 p_2 p_3, \dots, p_{\alpha\beta\gamma} = q_1 q_2 q_3$  in which the cellular probabilities can be written it is clear that the significance in question has to do with the complete independence or not of the three characters in the universe.<sup>4</sup> Had the assumption of complete independence been made in the first place it would have been seen that (1) was independent of  $p_1, p_2, p_3$  and that (3) followed as the value of the relative frequency.<sup>5</sup>

2. *One Marginal Face and the Complementary Edge Used.*—Of the six numbers  $AB, \alpha B, A\beta, \alpha\beta, C, \gamma$  only five are independent as the sum of the first four is equal to that of the last two and is  $N$ . There are three degrees of freedom in the sense that in terms of the values<sup>6</sup> of  $ABC, AC, BC$ , one has  $AB\gamma = AB - ABC$  and

$$\begin{aligned} \alpha BC &= BC - ABC, A\beta C = AC - ABC, \alpha\beta C = C - AC - BC + ABC, \\ \alpha B\gamma &= \alpha B - BC + ABC, A\beta\gamma = A\beta - AC + ABC, \\ \alpha\beta\gamma &= \alpha\beta - C + AC + BC - ABC. \end{aligned}$$

It thereupon turns out that  $p_{ABC} = p_{11}p$ ,  $p_{AB\gamma} = p_{11}q$ ,  $\dots$ ,  $p_{\alpha\beta\gamma} = p_{22}q$  in terms of four unknowns  $p_{11}, p_{12}, p_{21}, p_{22}$  whose sum is 1 and an additional unknown  $p$ . The value of the relative probability of a table is

$$\frac{AB! \alpha B! A\beta! \alpha\beta! C! \gamma!}{N! ABC! \dots \alpha\beta\gamma!} \quad (4)$$

The problem is to discuss whether the samples of  $C$  and of  $\gamma$  come from the same  $2 \times 2$  universe<sup>7</sup> of  $A$  and  $B$ . The expected value of  $(ABC)$  is  $Np_{11}p$  and is unknown, but if  $p_{11}$  and  $p$  be estimated from the marginal data it would be  $(ABC) = (AB)(C)/N$  and similarly for the other tabular elements, from which  $\chi^2$  could then be computed.

3. *Two Marginal Faces Used.*—Of the eight elements in the two marginal faces only six are independent; they may be taken as  $AB$ ,  $\alpha B$ ,  $A\beta$ ,  $\alpha\beta$ ,  $AC$ ,  $\alpha C$ , leaving two degrees of freedom to be specified by  $ABC$  and  $BC$ . As in the first case eight unknown values of  $p$  could be expressed in terms of three independent unknown probabilities and in the second case in terms of four, so here they can be expressed in terms of five. Indeed, the part of  $P$  in (1) that depends on the probabilities in the universe may be written

$$p_{ABC}^{ABC} p_{AB\gamma}^{AB} - {}^{ABC}p_{A\beta C}^{AC} - {}^{ABC}p_{\alpha B C}^{BC} - {}^{ABC}p_{A\beta\gamma}^{A} - {}^{AB} - {}^{AC} + {}^{ABC} \times \\ p_{\alpha B\gamma}^{B} - {}^{AB} - {}^{BC} + {}^{ABC} p_{\alpha\beta C}^{C} - {}^{AC} - {}^{BC} + {}^{ABC} p_{\alpha\beta\gamma}^{N} - {}^A - {}^B - {}^C + {}^{AB} + {}^{AC} + {}^{BC} - {}^{ABC} \\ \text{or } \left( \frac{p_{ABC} p_{AB\gamma} p_{\alpha B\gamma} p_{\alpha\beta C}}{p_{\alpha\beta\gamma} p_{\alpha\beta C} p_{A\beta C} p_{AB\gamma}} \right)^{ABC} \left( \frac{p_{\alpha\beta C} p_{\alpha\beta\gamma}}{p_{\alpha B\gamma} p_{\alpha\beta C}} \right)^{BC} \times \\ p_{ABC}^{AB} p_{A\beta C}^{AC} p_{A\beta\gamma}^{\alpha B} p_{\alpha B\gamma}^{A\beta} p_{\alpha\beta C}^{\alpha C} p_{\alpha\beta\gamma}^{\alpha} - {}^{\alpha B} - {}^{\alpha C},$$

whence

$$\frac{p_{\alpha B C} p_{\alpha\beta\gamma}}{p_{\alpha B\gamma} p_{\alpha\beta C}} = \frac{p_{A\beta C} p_{AB\gamma}}{p_{ABC} p_{A\beta\gamma}} = 1.$$

These equations are the conditions that  $B$  and  $C$  are not associated in the subuniverses of  $\alpha$  and of  $A$ . From this it follows that

$$p_{ABC} = \frac{p_{AB} p_{AC}}{p_A}, \quad p_{\alpha BC} = \frac{p_{\alpha B} p_{\alpha C}}{p_\alpha}, \quad \dots, \quad p_{\alpha\beta\gamma} = \frac{p_{\alpha\beta} p_{\alpha\gamma}}{p_\alpha}.$$

The chance of getting the partition  $AB$ ,  $\alpha B$ ,  $A\beta$ ,  $\alpha\beta$  from  $N$  and the partitions  $AC$ ,  $A\gamma$  from  $A$  and  $\alpha C$ ,  $\alpha\gamma$  from  $\alpha$  is

$$\frac{N! \quad A! \quad \alpha!}{AB! \alpha B! A\beta! \alpha\beta! AC! A\gamma! \alpha C! \alpha\gamma!}$$

and hence the relative frequencies are

$$\frac{AB! \alpha B! A\beta! \alpha\beta! AC! A\gamma! \alpha C! \alpha\gamma!}{A! \alpha! ABC! \dots \alpha\beta\gamma!} \quad (5)$$



The problem here is whether the observed  $2 \times 2 \times 2$  table could have arisen from two different non-associated universes,<sup>8</sup> one in  $A$  and the other in  $\alpha$ .

4. *The Three Faces Are Used.*—Here there is only one degree of freedom which may be specified by  $ABC$ ; there are seven independent unknown frequencies. The condition for independence of  $P$  of the unknown  $p$ 's is readily shown to be

$$\frac{p_{ABC}p_{AB\gamma}p_{\alpha B\gamma}p_{\alpha\beta C}}{p_{\alpha\beta\gamma}p_{AB\gamma}p_{\alpha BC}p_{ABC}} = 1 \text{ or } \frac{p_{ABC}p_{AB\gamma}}{p_{AB\gamma}p_{ABC}} = \frac{p_{\alpha BC}p_{\alpha\beta\gamma}}{p_{\alpha\beta\gamma}p_{\alpha B\gamma}}.$$

The association coefficients of Yule, namely,

$$Q = \frac{AB \cdot \alpha\beta - \alpha B \cdot A\beta}{AB \cdot \alpha\beta + \alpha B \cdot A\beta} \text{ and } \omega = \frac{\sqrt{AB}\sqrt{\alpha\beta} - \sqrt{\alpha B}\sqrt{A\beta}}{\sqrt{AB}\sqrt{\alpha\beta} + \sqrt{\alpha B}\sqrt{A\beta}}$$

are both functions of the quotient  $AB \cdot \alpha\beta / \alpha B \cdot A\beta$ . The condition just found for the  $p$ 's shows that  $B$  and  $C$  have the same association coefficients in the subuniverses of  $A$  and  $\alpha$  (and because of symmetry  $A$  and  $C$  have the same association coefficients in the subuniverses of  $B$  and  $\beta$ , and so for  $B$  and  $C$  in  $A$  and  $\alpha$ ). The problem is to determine whether the observed table could reasonably have arisen from such a  $2 \times 2 \times 2$  universe.<sup>9</sup> We have not yet been able to find a formula for the relative probabilities of all the tables in the linear series but when numbers are small the reciprocal of the product of the factorials of the cell frequencies of all tables of the series may be tabulated and the relative frequencies found.<sup>10</sup>

Although the value of the probabilities in the universe must remain unknown as in previous cases because the very essence of this approach is that the results must be independent of those unknown probabilities, one may assign values to them in terms of the marginal elements that are used, and these marginal elements would define the probabilities if the sampling were really from a universe in which the marginal elements used were really fixed by external constraints instead of appearing by virtue of the sampling process. For the present case the expected value  $Np_{ABC}$  or  $ABC$  would then be the value of  $x$  which satisfies the cubic equation

$$(ABC - x)(A\beta\gamma - x)(\alpha B\gamma - x)(\alpha\beta C - x) = (\alpha\beta\gamma + x)(AB\gamma + x)(\alpha BC + x)(A\beta C + x)$$

or the value of  $p_{ABC}$  would come from the cubic equation

$$p_{ABC}^3 - p_{ABC}^2[p_{AB} + p_{AC} + p_{BC} - p_A p_B - p_B p_C - p_A p_C + p_B p_{AC} + p_A p_{BC} + p_C p_{AB}] + p_{ABC}[2p_{AC}p_{AB}p_{BC} + p_{AB}p_{AC} + p_{AB}p_{BC} + p_{AC}p_{BC} + p_B p_{AC}^2 + p_A p_{BC}^2 + p_C p_{AB}^2 + p_A p_B p_C - p_A p_C p_{AB} - p_A p_C p_{BC} - p_B p_C p_{AB} - p_B p_C p_{AC} - p_A p_B p_{AC} - p_A p_B p_{BC}] - p_{AB}p_{AC}p_{BC}[1 - p_A - p_B - p_C + p_{AB} + p_{AC} + p_{BC}] = 0.$$

The value of  $x$  must lie between the least of the four frequencies  $ABC$ ,  $A\beta\gamma$ ,  $\alpha B\gamma$ ,  $\alpha\beta C$  and the negative of the least of the four frequencies  $\alpha\beta\gamma$ ,  $AB\gamma$ ,  $A\beta C$ ,  $\alpha BC$ . It may be observed that the expected value is here not the mean value<sup>1</sup> in the series as it is in the previous three cases and in the  $m \times n$  contingency table.

<sup>1</sup> See, these PROCEEDINGS, 28, 378-384 (1942).  
<sup>2</sup> For example, if the table be  $ABC = 0$ ,  $AB\gamma = 0$ ,  $A\beta C = 0$ ,  $\alpha BC = 0$ ,  $A\beta\gamma = 1$ ,  $\alpha B\gamma = 2$ ,  $\alpha\beta C = 3$ ,  $\alpha\beta\gamma = 1$  there are 10 tables consistent with  $N = 7$ ,  $A = 1$ ,  $B = 2$ ,  $C = 3$ , viz.,

	$ABC$	$AB\gamma$	$A\beta C$	$\alpha BC$	$A\beta\gamma$	$\alpha B\gamma$	$\alpha\beta C$	$\alpha\beta\gamma$
1	0	0	0	0	1	2	3	1
2	0	0	0	1	1	1	2	2
3	0	0	0	2	1	0	1	3
4	0	0	1	0	0	2	2	2
5	0	0	1	1	0	1	1	3
6	0	0	1	2	0	0	0	4
7	0	1	0	0	0	1	3	2
8	0	1	0	1	0	0	2	3
9	1	0	0	0	0	1	2	3
10	1	0	0	1	0	0	1	4

<sup>3</sup> The relative probabilities of the above tables are  
1        2        3        4        5        6        7        8        9        10  
4/49   12/49   4/49   6/49   8/49   1/49   4/49   4/49   4/49   2/49  
Only set-up 6 is significant at the 0.05 level with a relative probability of 1/49.

<sup>4</sup> If it were desired to use the marginal values to estimate  $p_1 = p_A$  as  $(A)/N$ , etc., the expected value of  $(ABC)$  would be  $(A)(B)(C)/N^2$ , etc., and then  $\chi^2$  could be computed.

<sup>5</sup> If there were  $k$  characters, the number of degrees of freedom for complete independence would be  $2^k - k - 1$  and the relative probability would be

$$\frac{A! \alpha! B! \beta! \dots K! \kappa!}{(N!)^{k-1} (AB \dots K)! \dots (\alpha\beta \dots \kappa)!}$$

Even with small numbers there would be serious difficulty in carrying out the arithmetic for a particular case.

<sup>6</sup> For example, let  $N = 6$ ,  $C = 4$ ,  $A = 3$ ,  $B = 3$ ,  $AB = 2$ . There are consistent with these boundary conditions 8 tables which are as follows:

	$ABC$	$AB\gamma$	$A\beta C$	$\alpha BC$	$A\beta\gamma$	$\alpha\beta\gamma$	$\alpha\beta C$	$\alpha\beta\gamma$
1	1	1	1	0	0	1	2	0
2	0	2	1	1	0	0	2	0
3	2	0	1	0	0	1	1	1
4	2	0	0	0	1	1	2	0
5	1	1	1	1	0	0	1	1
6	1	1	0	1	1	0	2	0
7	2	0	1	1	0	0	0	2
8	2	0	0	1	1	0	1	1

The relative probabilities of these tables are

1	2	3	4	5	6	7	8
2/15	1/15	2/15	1/15	4/15	2/15	1/15	2/15

<sup>7</sup> The natural generalization would be to consider whether  $k$  tables containing  $C_1, C_2, \dots, C_k$  elements could arise from the same  $2 \times 2$  universe. The relative probabilities of the different  $2 \times 2 \times k$  tables would be

$$\frac{AB! \alpha B! A\beta! \alpha\beta! C_1! C_2! \dots C_k!}{N! \times \text{product of factorials of cell frequencies}}$$

<sup>8</sup> The natural generalization would be to consider whether a  $2 \times 2 \times k$  table could have originated from  $k$  non-associated universes by taking samples of  $A_1, A_2, \dots, A_k$  from them and the relative probabilities would be

$$\frac{A_1B! A_1\beta! A_1C! A_1\gamma! \dots A_kB! A_k\beta! A_kC! A_k\gamma!}{A_1! \alpha_1! \dots A_k! \alpha_k! \times \text{product of factorials of cell frequencies}}$$

<sup>9</sup> The natural generalization would be to  $k$  characters and a  $2^k$  table. The one condition on the  $2^k$  values of the cell probabilities in the universe would be that the product of the  $2^{k-1}$  values of  $p$  for cells corresponding to  $k, k-2, k-4, \dots$  positive characters would have to be equal to the product of the  $2^{k-1}$  values of  $p$  for cells corresponding to  $k-1, k-3, k-5, \dots$  positive characters. It is to be understood that all combinations of  $k-1$  characters are used.

<sup>10</sup> If  $N = 45, A = 19, B = 14, C = 20, AB = 5, AC = 8, BC = 7$  there are 6 tables consistent with these boundary conditions which are as follows:

	$ABC$	$AB\gamma$	$A\beta C$	$\alpha BC$	$A\beta\gamma$	$\alpha B\gamma$	$\alpha\beta C$	$\alpha\beta\gamma$
1	5	0	3	2	11	7	10	7
2	4	1	4	3	10	6	9	8
3	3	2	5	4	9	5	8	9
4	2	3	6	5	8	4	7	10
5	1	4	7	6	7	3	6	11
6	0	5	8	7	6	2	5	12

The relative frequencies of these tables as determined from the reciprocal of the product of the factorials of the cell frequencies are

1	2	3	4	5	6
48	1925	11550	13860	3360	126
<u>30869</u>	<u>30869</u>	<u>30869</u>	<u>30869</u>	<u>30869</u>	<u>30869</u>

Numbers 1 and 6 are significant at the 0.05 level.

<sup>11</sup> Note that the mean value of  $ABC$  in the above series is 2.3865 whereas the expected value as determined from the cubic equation is 2.3884.

# PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

Volume 28

October 15, 1942

Number 10

Copyright 1942 by the National Academy of Sciences

## *A COMPLEMENT-RELEASE REACTION; THE NEUTRALIZATION OF THE ANTICOMPLEMENTARY ACTION OF SEA-URCHIN FERTILIZIN BY ANTIFERTILIZIN\**

BY ALBERT TYLER

WILLIAM G. KERCKHOFF LABORATORIES OF THE BIOLOGICAL SCIENCES, CALIFORNIA  
INSTITUTE OF TECHNOLOGY

Communicated August 23, 1942

In previous articles<sup>1-8</sup> the interaction of fertilizin and antifertilizin has been considered analogous to that of an antigen and its specific antibody in the more usual serological reactions. Since many antigen-antibody reactions exhibit complement-fixation, it was of interest to learn if this occurs in the union of fertilizin and antifertilizin. A positive result would strengthen the analogy with serological reactions. However, it is by no means a necessary characteristic of antigen-antibody reactions that they fix complement *in vitro*, since there are a great many exceptions, such as all toxin-antitoxin reactions that have been examined,<sup>9</sup> the reaction of horse<sup>10</sup> or human<sup>11</sup> antipneumococcus antibodies with capsular carbohydrate, and most cases of the action of the normal iso-antibodies of the blood-groups.<sup>12</sup> Also, in many instances antigen or antibody alone are too highly anticomplementary to permit a determination of complement fixation by their interaction.

In complement-fixation tests with sea-urchin fertilizin-antifertilizin mixtures, the fertilizin was found to be highly anticomplementary. Dilutions below the anticomplementary limit gave with antifertilizin no complement-fixation. Above the anticomplementary limit a new phenomenon was found; namely, that the addition of antifertilizin neutralized the anticomplementary action of fertilizin, or if the complement were first bound by fertilizin, antifertilizin released it. This phenomenon is therefore designated a "Complement-Release Reaction." The data show it to be a quantitative measure of the interaction of fertilizin and anti-fertilizin and to involve the fourth component (C'4) of complement.

*Material and Methods.*—The fertilizin and the antifertilizin solutions were partially purified preparations from the eggs and the sperm, respectively, of the sea-urchin *Strongylocentrotus purpuratus*, prepared according

TABLE I  
NEUTRALIZATION OF ANTICOMPLEMENTARY ACTION OF *S. purpuratus* FERTILIZIN BY ANTIFERTILIZIN

Two units of complement (0.25 ml. of 1/25 guinea-pig serum) added to indicated mixtures, incubated 1 hour at 37°C., sheep cells added and degree of hemolysis read after 1 hour at 37°C.

++++ = Complete hemolysis; 0 = no hemolysis

Dilutions of 0.25 ml. of ADJUSTED FERTILIZIN (TITER = 512)	DILUTIONS OF 0.25 ML. OF ADJUSTED FERTILIZIN (TITER = 512)								RINGERS	NO COMPLEMENT
	1	1/2	1/4	1/8	1/16	1/32	1/64	1/128		
1	++	++	++	++	++	++	++	++	++	0
1/2	++	++	++	++	++	++	++	++	++	0
1/4	++	++	++	++	++	++	++	++	++	0
1/8	++	++	++	++	++	++	++	++	++	0
1/16	++	++	++	++	++	++	++	++	++	0
1/32	++	++	++	++	++	++	++	++	++	0
1/64	++	++	++	++	++	++	++	++	++	0
1/128	++	++	++	++	++	++	++	++	++	0
1/256	++	++	++	++	++	++	++	++	++	0
Ringers	++	++	++	++	++	++	++	++	++	0

to previously described methods. Fertilizin titer is given as the reciprocal of the final dilution that gives microscopically perceptible agglutination with an equal volume of a 1% sperm suspension. Antifertilizin titer is the reciprocal of the dilution that just neutralizes one unit of fertilizin. For use in the hemolytic system fertilizin and antifertilizin solutions were adjusted to 0.9% salinity. Pooled fresh guinea-pig serum was used as the source of complement. The indicator consisted of a 2% suspension of sheep erythrocytes sensitized with three units of amboceptor (rabbit antisheep cell serum) and 0.5 ml. of this was used in a total reactant volume of 1.25 ml. All dilutions and adjustments to total volume were made with a "sea water-Ringers" solution prepared by diluting three volumes of sea water (local salinity = 3.3%) with eight volumes of distilled water.

Interaction of Fertilizin, Antifertilizin and Complement.—Solutions of *S. purpuratus* fertilizin are found to be very highly anticomplementary and the extent of this action roughly parallels the sperm-agglu-

tinating titer. In the experiment listed in table 1 (last line) 0.25 ml. of a 32 unit fertilizin solution ( $1/16$  dilution) inactivates 2 units of complement. In 13 other experiments the inactivation ratio ranged from 8:1 to 32:1 with an average of 20:1. On the other hand concentrated solutions of antifertilizin show no anticomplementary action (next to last column of table) nor do they have any hemolytic action on sensitized sheep cells (last column of table). The addition of antifertilizin overcomes the anticomplementary action of fertilizin and, as the data show, the amount required is directly proportional to the amount of fertilizin present. In these mixtures one might have expected to find complement fixation. Instead the presence of antifertilizin protects complement from inactivation by fertilizin. There is no sign of complement fixation with dilutions (such as  $1/128$  in the table) of fertilizin that are below the anticomplementary limit. From the data in the table it may be seen that one unit of antifertilizin neutralizes the anticomplementary action of 2 to 4 units of fertilizin and combined with 13 other less extensive experiments the ratio approximates 1:4.

If complement is allowed to react with fertilizin for some time before the addition of antifertilizin the results are the same as when all three are mixed immediately. If the sensitized sheep cells are added directly after the addition of antifertilizin, the same degree of hemolysis is obtained but there is a definite time lag when compared with a previously incubated mixture. Since antifertilizin has no hemolytic action on sensitized sheep cells the results mean that complement as a whole is not destroyed or irreversibly bound by fertilizin.

The interaction of fertilizin and complement is also manifested by an inhibition of the sperm-agglutinating power of the former. For these tests the guinea-pig serum is adjusted to sea water salinity. In four experiments that were run one unit of complement neutralized the sperm-agglutinating action of 4 to 7 (av. 5) units of fertilizin. This is much less than might be expected from the anticomplementary ratio (20:1), but it is consistent with the ratios of fertilizin to antifertilizin that remove the anticomplementary action of the former. It may be interpreted as signifying a stronger interaction of fertilizin with antifertilizin (either in solution or on the sperm) than with complement or, in other words, a lower equilibrium constant for the former reaction than for the latter.

When guinea-pig serum is added to intact *S. purpuratus* eggs the latter agglutinate and a precipitation membrane (<sup>4</sup> fig. 1) forms on the gelatinous coat. Since the jelly coat is composed of fertilizin<sup>5,6</sup> this action is, very likely, simply another manifestation of its interaction with complement.

*Action of Univalent Fertilizin.*—It has been shown<sup>6,12</sup> that fertilizin can be converted into a non-agglutinating form, termed "univalent," that still combines with the species sperm or with antifertilizin. This, then, offered another means for testing for complement-fixation, if the anticomplemen-

tariness of fertilizin were lost by the conversion. However, in three experiments that were run the univalent fertilizin was found to be just as anticomplementary as the unaltered agent. In addition its effect on complement was likewise found to be neutralized by antifertilizin.

*Component of Complement Involved.*—Complement is known to be composed of two heat labile and two relatively heat stable components, now designated,<sup>14,15</sup> respectively, C'1 (mid-piece), C'2 (end-piece), C'3 (third component) and C'4 (fourth component). Guinea-pig serum that had been heated at 56°C. for 1½ hour to inactivate C'1 and C'2 was found to be capable of reactivating complement that had been inactivated by fertilizin. Also, such heated serum neutralized the sperm-agglutinating power of fertilizin and agglutinated the intact sea-urchin eggs. Serum absorbed with yeast (° p. 16) to remove C'3 was likewise found to be capable of reactivating fertilizin-inactivated complement and agglutinating the eggs, while serum treated with ammonia<sup>16</sup> to inactivate C'4 lacked the ability. These results are illustrated in table 2 which contains the data of one of four similar experiments. It is clear, then, that fertilizin combines with C'4.

*Failure of Antifertilizin to Act as C'4.*—A reasonable interpretation of the fact that antifertilizin and C'4 can displace one another from combination with fertilizin is that both possess very similar combining groups. On this basis one would expect antifertilizin to be capable of acting as a substitute for C'4 in the hemolytic system. Attempts were therefore made to reactivate ammonia-treated guinea-pig serum by the addition of antifertilizin. The results (table 2) were consistently negative, however. This might mean that while antifertilizin resembles C'4 in some respects

TABLE 2  
REACTIVATION OF FERTILIZIN-INACTIVATED COMPLEMENT AND TEST OF ANTIFERTILIZIN FOR C'4 ACTIVITY

DILUTIONS OF 0.25 ML. OF:					DEGREE OF HEMOLYSIS
FERTILIZIN (TITER = 256)	NORMAL GUINEA-PIG SERUM	YEAST- ABSORBED GUINEA-PIG SERUM	AMMON-TREATED GUINEA-PIG SERUM	ANTIFERTILIZIN (TITER = 128)	
...	1/25	...	.....	.....	++++
1/16	1/25	...	.....	.....	0
1/32	1/25	...	.....	.....	+
1/64	1/25	...	.....	.....	+++
1/128	1/25	...	.....	.....	++++
...	...	1/2	.....	.....	0
...	...	...	1/2	.....	0
...	...	1/20	1/20	.....	++++
1/16	1/25	1/20	.....	.....	++++
1/16	1/25	...	1/2	.....	0
...	...	...	1/2 and less	1 and less	0

it differs in properties that enable C'4 to collaborate with the other three components and the sensitized cells in the hemolytic system. Another interpretation is that the mutual displacement is due, not to similarity of combining groups of antifertilizin and C'4, but to sufficiently close proximity of combining sites on the fertilizin molecule so that union with one of them inhibits union with the other.

*Summary.*—The fertilizin obtained from sea-urchin eggs is found to be highly anticomplementary. There is no evidence for complement-fixation in its interaction with antifertilizin from the species sperm, but instead there appears a new type of phenomenon which is termed "Complement Release." This consists in the liberation by antifertilizin of complement that has been bound by fertilizin or in the neutralization by antifertilizin of the anticomplementary action of fertilizin. The effect is a quantitative measure of the interaction of fertilizin and antifertilizin. Only the fourth component (C'4) of complement is involved and the results may imply chemical similarity between C'4 and antifertilizin. However, antifertilizin cannot replace C'4 in the hemolytic system.

I am indebted to Professor Sterling Emerson for his generous coöperation and advice in this work.

\* This investigation was aided by a grant from the Penrose Fund of the American Philosophical Society.

<sup>1</sup> Tyler, A., *Proc. Nat. Acad. Sci.*, **25**, 317 (1939).

<sup>2</sup> Tyler, A., and Fox, S. W., *Science*, **90**, 516 (1939).

<sup>3</sup> Tyler, A., *Biol. Bull.*, **78**, 159 (1940).

<sup>4</sup> Tyler, A., *Proc. Nat. Acad. Sci.*, **26**, 249 (1940).

<sup>5</sup> Tyler, A., and Fox, S. W., *Biol. Bull.*, **79**, 153 (1940).

<sup>6</sup> Tyler, A., *Ibid.*, **81**, 190 (1941).

<sup>7</sup> Tyler, A., and O'Melveny, K., *Ibid.*, **81**, 364 (1941).

<sup>8</sup> Tyler, A., *West. Jour. Surg., Obstet. & Gynec.*, **50**, 126 (1942).

<sup>9</sup> Osborn, T. W. B., *Complement or Alexin*, Oxford Univ. Press, 1937, p. 36.

<sup>10</sup> Zinsser, H., and Parker, J. J., *Jour. Exp. Med.*, **49**, 251 (1923).

<sup>11</sup> Stats, D., and Bulowa, J. G. M., *Jour. Immunol.*, **44**, 41 (1942).

<sup>12</sup> Wiener, A. S., *Blood Groups and Blood Transfusion*, Charles C. Thomas, Baltimore, 1939, p. 25.

<sup>13</sup> Metz, C. B., *Biol. Bull.*, **82**, 446 (1942).

<sup>14</sup> Pillemer, L., and Ecker, E. E., *Science*, **94**, 437 (1941); *Ann. N. Y. Acad. Sci.*, **43**, 63 (1942).

<sup>15</sup> Heidelberger, M., *Jour. Exp. Med.*, **73**, 681 (1941).

<sup>16</sup> Whitehead, H. R., Gordon, J., and Wormall, A., *Biochem. Jour.*, **19**, 618 (1925).



WHITE PIGMENTARY EFFECTORS (LEUCOPHORES) IN  
KILLIFISHES

By E. F. B. FRIES

DEPARTMENT OF BIOLOGY, THE CITY COLLEGE, NEW YORK

Communicated September 1, 1942

In interpreting their observations on birefringent materials in the scales of *Fundulus heteroclitus* examined with polarized light, Shanes and Nigrelli (1941) seem to deny the morphological individuality of what Odiorne (1933) called guanophores. After first identifying the birefringent masses they observed as the guanophores of Odiorne, they advocate the conclusion that the doubly refractive material is an integral part of the scale melanophores and xanthophores. The question thus raised as to the status of the guanophores, about which relatively little is known (Parker, 1940), may lend heightened interest to observations of mine on certain similar integumentary structures found in other, related species. In any event, these observations add to available pertinent information concerning species found very useful for study of the complexities of the hormonal and nervous control of their often independently responding different kinds of pigmentary effectors.

While a guest in the U. S. Fish and Wildlife Laboratory at Beaufort, N. C.—where I was hospitably and helpfully accorded all facilities by the Director, Dr. H. F. Prytherch—I had occasion to study the chromatophores of the dorsal aspect of local Cyprinodontidae, especially the striped killifish, *Fundulus majalis* Wahlb. Very numerous, chromatophorelike, whitish structures at once attracted my attention.

The component material that made these bodies conspicuous in *F. majalis* was not iridescent and appeared to share actively in chromatic adjustments; also, they occurred in addition to iridescent bodies as well as melanophores and xanthophores. They resembled published photographic representations of the "guaninophores" (Ginsburg, 1929) and silvery halolike "iridosomes" (Sumner and Wells, 1933) of *Lebistes*, which is classed in the same order as *Fundulus*, and were evidently the same kind of chromatophore as these and as the ones Odiorne described in *F. heteroclitus*.

Odiorne adhered to long-established views in classing the non-iridescent structures as guanophores. His conclusion that the minute, not obviously crystalline particles characterizing their contents was probably guanine rested on comparative grounds laid down in papers too numerous to relate (but traceable through the bibliographic citations given by Odiorne, Sumner and Wells, Ginsburg, and Foster). For my part, I did not determine by chemical means whether the whitish material in the structures that are the subject of the present paper was guanine. But there seems

to be no reason to doubt that it is the same as one of the forms in which MacMunn, for example, by his extensive chemical and cytochemical tests, identified guanine in divers teleosts (Cunningham and MacMunn, 1893). Modern analytic methods have not led to revision of the older determinations of guanine in different integumentary granules, etc., of whitish, silvery or iridescent appearance (Peschen, 1939; cf. also Millot, 1923).

Although the non-iridescent chromatophores in question would accordingly be guanophores, it seems to me preferable to call such bodies *leucophores*, i.e., white or colorless chromatophores. This designation, first used, it seems, by Keller (1895) in reference to similar structures in the chameleon, usefully indicates their distinctive chromatic character and does so without further (and perhaps rash) implications. It is more logically alternative to *iridocyte*, *iridophore* or *iridosome* than is *guanophore*, since the iridescent material of the former is just as probably guanine (in crystalline form) and justifies classing them as another type of guanophore, as Odiorne himself did. In construction and meaning, *leucophore* corresponds advantageously to *xanthophore*, *erythrophore*, etc., terms that conveniently distinguish chromatophores according to their color without reference to the chemistry of their pigments, which is not necessarily the same for chromatophores of one color.

Unlike *F. heteroclitus*, in which the "guanophores" occur only by rare and isolated exception in the outer dermal layer with its rich studding of melanophores, *F. majalis* is abundantly equipped with these white chromatophores, or leucophores, in the outer dermal layer over the dorsal and dorsolateral aspects of the whole body excepting the fins (into which few stray). Both species have also iridophores, commonly combined as iridosomes (which, like Odiorne, I distinguish from the "iridosomes" that Sumner and Wells record for *Lebistes*). The leucophores of *F. majalis* are characteristically, at least in the outer layer, closely associated with melanophores and melaniridosomes (described for *F. heteroclitus* by Foster, 1937); their central masses lie next internal to and usually hidden by the associated pigmentary bodies. Most of these outer chromatophores and chromatophore complexes are aggregated in diagonal bands crisscrossing the back and sides much the same as in *Lebistes* (cf. photographs of Sumner and Wells); they are grouped in less geometric pattern over the head. Intervening gaps of skin are rather transparent as deep as the inner dermal chromatophore layer beneath the scales, which is characterized by many more massive iridosomes or clusters of iridocytes and often larger and less delicately branched melanophores than those of the outer layer. (Superficial to all the other chromatophores is a sparse sprinkling of mostly smaller, simple melanophores. The relatively small but numerous xanthophores occurring both within the inner layer and just under the main outer dermal layer of other chromatophores, are fairly accessible to view

in the gaps of the latter.) Exceptionally, a leucophore of the outer dermal layer may occur apart from any melanophore or melaniridosome. The existence and mode of occurrence of these teleostean white chromatophores recalls certain Crustacea described as having polychromatic chromatophores including a white component and other species having white pigment in separate, or monochromatic, chromatophores (for bibliography, see Parker's review, 1940).

To determine whether the leucophores, combined or separate, do in fact play an active rôle in color changes shown macroscopically by the striped killifish, I tested them for response to background, i.e., to bottoms of different shades and hues. These were provided by 8-inch and 10-inch glass culture bowls painted externally with enamel (black, white, yellow or light blue). I enclosed each fish for observation in a shortened test tube opened to admit inflow of water at the head end and half-stoppered to permit outflow at the opposite end (Butcher, 1939). With the tube held in place, immersed in sea water in a painted bowl, by a metal strap pinched on the end away from the fish's eyes, the dorsal chromatophores could be examined through a binocular dissecting microscope, while the fish stayed in a situation inducing continued chromatic response to the given color. The bowls stood inside a window exposing the fish to a broad expanse of diffuse daylight. A Spencer "universal" microscope lamp provided good illumination of the microscopic field from above, without interfering detectably with the pigmentary adjustments called forth by the colored bowls.

The critical tests involved two male and four female *F. majalis* about 7 cm. long. In each, a group of favorably exposed chromatophores was selected for ready recognition and repeated inspection as the fish stayed now in one, now in an oppositely colored bowl. Such exchanges between black and white environments established the fact that the leucophores concentrated their "pigment" in response to the black and dispersed it in response to the white. After a stay of two hours or longer over the white bottom, the visible contents of every leucophore was spread finely through a delicate lacework of processes extending from the leucophore's center; they partly overlay the now concentrated pigment of the melanophores, obscuring them and the other chromatophores. When the fish was transferred to a black dish, the black pigment spread out in a few seconds part-way into the melanophore processes, thereby covering up much (but not yet the farthest processes) of any associated leucophores. The gray-whiteness of the slenderest and most peripheral lacy processes of each leucophore disappeared almost simultaneously. The means of this disappearance could not be the screening effect of the melanin, which was not yet so widely dispersed. As the melanophores continued showing rapid pigment spread, a markedly slower centripetal accumulation of the white

material was discernible in the leucophores (usually increasingly obscured by the melanophores, but with substantial opaquely white stumps often to be seen between the roots of the melanophore processes). After longer sojourn over black, the melanophores in the maximum pigment dispersal never duplicated the gossamerlike appearance characterizing the leucophores in their maximum dispersal. The extreme dispersed form of a melanophore was not identical with that of its associated leucophore. Evidently the processes of the two were complexly interlaced, not confluent. Upon reverse transfers of the fully black-adapted fish to a white or other pale bowl, a great reduction of melanin dispersal occurred in a few seconds, whereas it took longer for the white processes to appear; and the melanin concentration seemed practically complete well before the dispersing response of the white chromatophores approximated its maximum. The leucophores took more than half an hour (at about 20°C.) to bring their dispersing change close to completion; this was twice as long as the melanophores required for equivalent pigment concentration. Whether, indeed, the leucophore changes consisted of such dispersing and concentrating migrations of particles outward and inward in the processes as are familiar in melanophores could not be seen with the magnifications used (higher than 3.4 $\times$  objectives and 12.5 $\times$  oculars proved impractical, especially because of the fish's breathing movements). In any case, the conclusion seems clear: the leucophores of this species are active effectors; they assist in the color changes whereby the fish becomes less conspicuous over different grounds and they do so by responding, inversely as compared with the quicker reacting melanophores, to the shade of the bottom.

Examination of the leucophores after similar sojourns over, and transfers between, yellow and blue revealed no sure differences. Comparative examination of six fish macroscopically well adapted to blue and a larger lot of others well adapted to yellow added support to the conclusion that leucophore changes are probably correlated only with the shade and not with the hue of the environment. There appeared at most a mere suggestion of more filmy, extreme, leucophore dispersal and complete screening of the black and iridescent chromatophores in the blue than in the yellow fish; such an apparent difference might be illusory, resulting from the condition of the other, deeper-lying pigments (the dorsal xanthophores showed decided pigment dispersal in the yellow, versus concentration in the blue, fish).

Leucophores are very plentiful, in addition to iridophores, also in the sheephead minnow, *Cyprinodon variegatus* Lacépède. They occur apart from melanophores less rarely than in *F. majalis*. This made it easier to determine that they quickly initiated dispersal in response to white and concentration in response to black. One of these fish, 3 cm. long, was tested through repeated stays in the black, white, yellow and blue bowls;

it showed no significant difference from *F. majalis* in the behavior of the leucophores. It was particularly in *Cyprinodon*, however, that a close relationship between iridophores and non-iridescent leucophores was indicated by a similarity of the reactive leucophores, in size and glitter of some of their contained bodies, to the same fish's iridophores, which were unreactive (constant in form).

Intraperitoneal injections into eight *F. majalis* of ergotamine tartrate (0.01–0.03 ml. of Sandoz's "gynergen" per gram of fish) demonstrated that the response of the leucophores can take place without the opposite changes occurring in the melanophores.\* While the latter remained in the concentrated state regardless of what vessel the fish was kept in, the leucophores still effected concentration in the black bowls and dispersal in the white. The demonstration was especially convincing in the case of two fish that had been hypophysectomized fifteen days before the ergotization. After a two-hour stay over black or white, following transfer from the opposite color, repeatedly in these fish the leucophores responded as already described, while the melanophores remained practically punctate. Assuredly, then, whatever the longer-term biochemical relation may be between the black and white chromatophores, and however intimately they are associated, the white chromatophores are not undissociable parts of other chromatophores, but chromatically individual functional entities, significant especially for the maximum pallor achieved in this species. Their resemblance to the independently reactive white pigmentary effectors of crustaceans accordingly embraces their function.

It follows from these results of ergotization that the mechanisms mediating the presumably visually initiated response of the leucophores to pale or dark ground colors must differ from those subserving the similarly adaptive changes of the melanophores and xanthophores. That the mechanisms exclude direct innervation of the leucophores suggests itself but is unproved by present data, since the ergot dose that blocks melanophore changes does not paralyze such nervous functions in the fish as equilibration, breathing, etc. On the other hand, it follows from the same experiments performed on pituitaryless fish (five of the eight that were ergotized) that the leucophores do not depend on variation in the supply of a pituitary hormone for their changes in response to the shade of the bottom.

*Summary.*—White chromatophores like the non-iridescent guanophores of *Fundulus heteroclitus*, but preferably called leucophores, are abundant in the outer derma of *F. majalis* and *Cyprinodon variegatus*.

They participate actively, but slower than the melanophores, in chromatic responses to "backgrounds," effecting concentration of their whitish contents in fish kept in a black bowl and presenting a very different, dispersed appearance, which augments pallor, in those kept over a white, light blue or yellow bottom.

Ergotization of *F. majalis* kept the melanophores in the concentrated state without stopping the leucophore changes. This confirmed that the leucophores are functionally individual, rather than integral parts of the melanophores with which most are combined.

They do not (in *F. majalis*) depend on the pituitary for mediation of their responses.

\* A fuller report of the effects of ergotamine on the several chromatophores in this species is in preparation for publication elsewhere.

Butcher, E. O., *Biol. Bull.*, **77**, 258-267 (1939).

Cunningham, J. T., and MacMunn, C. A., *Phil. Trans. Roy. Soc. (London)*, **B184**, 765-812 (1893).

Foster, K. W., *Jour. Exp. Zool.*, **77**, 169-213 (1937).

Ginsburg, J., *Zool. Jahrb., Abt. Anat.*, **51**, 227-280 (1929).

Keller, R., *Arch. ges. Physiol.*, **61**, 123-168 (1895).

Millot, J., *Bull. Biol. France Belg.*, **57**, 261-363 (1923).

Odiorne, J. M., *Proc. Nat. Acad. Sci.*, **19**, 750-754 (1933).

Parker, G. H., *Proc. Amer. Acad. Arts Sci.*, **73**, 165-195 (1940).

Peschen, K. E., *Zool. Jahrb., Abt. Physiol.*, **59**, 429-462 (1939).

Shanes, A. M., and Nigrelli, R. F., *Zoologica (N. Y. Zool. Soc.)*, **26**, 237-243 (1941).

Sumner, F. B., and Wells, N. A., *Jour. Exp. Zool.*, **64**, 377-403 (1933).

---

## DISTORTION OF STRATIGRAPHIC THICKNESSES DUE TO FOLDING

BY ERNST CLOOS

DEPARTMENT OF GEOLOGY, THE JOHNS HOPKINS UNIVERSITY

Communicated August 22, 1942

*Introduction.*—Discussions of causes or mechanism of folding include as one vital component the thicknesses of the folded strata. Geosynclines, troughs or furrows are regions in which sedimentation has led to greater stratigraphic thicknesses and out of which folded zones, like the Appalachians, emerge. Thousands of papers have been written with the assumption that thicknesses as measured today are indicative of and permit conclusions of depths of troughs, basins of sedimentation, location of geosynclines, correlation of increase or decrease of thicknesses of formations and many others. Swells and deeps within geosynclines have been discussed and Schuchert's paleogeographic maps are well known to every geologist. The concept of the geosyncline has become one of the pillars on which tectonic speculation rests. The author, however, feels obliged to cast some serious doubts on the underlying assumptions which are contained in the determination of stratigraphic thicknesses.

The general assumption may be phrased simply as follows: "Folding



(of the Appalachian or Alpine type) occurs in troughs or geosynclines where sediments are thick." But the question may also be put as follows: "Sediments are thick because they are folded." Little attention has been paid to the latter possibility and nobody as yet has eliminated it definitely. The author realizes, of course, that lateral shortening leads to thickening of the crust. Thickening of beds in anticlinal crests has long been recognized but the question arises whether or not thicknesses as measured in folded regions are thicknesses of deposition. In the following the author tries to show that "established" sections may be incorrect and that much of our knowledge of geosynclines may be based on alarmingly scant evidence.

*Method of Investigation.*—If stratigraphic thicknesses as measured today do not indicate thickness of sedimentation or depths of basins, they must be altered secondarily during folding. This leads to an investigation of deforming processes and quantitative determinations of deformation. The second step in the analysis consists of a comparison of measured thicknesses as described in the literature with known deformational distortion. Incidental to this phase is a critical analysis of some of the better known recorded thicknesses.

The author cannot possibly treat the enormous literature comprehensively here and only wishes to call attention to some of the principles involved. He feels justified in presenting the data because the problem is vital and far-reaching and the consequences should be called to the attention of the field geologist who deals with folded sediments.

The discussion is restricted to a rather small area in the Appalachians but applies everywhere. Quantitative measurements of the deformation of oolites were undertaken in Maryland and Pennsylvania in Cambro-Ordovician limestones under a grant by the Geological Society of America. A detailed and comprehensive account on that work will follow later.

*Folding.*—Folding is meant to include the formation of folds of any kind, the bending of strata into anticlines and synclines whether symmetrical, overturned, recumbent, parallel or isoclinal. It includes the formation of a flow cleavage more or less parallel to the axial plane and relative movements on these planes. Shear folds and rock flowage as shown in thickening and thinning of beds within portions of folds is included. Deformation by fracturing, faulting, jointing, etc., are excluded for the present purpose.

Inasmuch as the distortion of stratigraphic thickness forms the main topic of discussion, such processes as may contribute to this distortion are included.

*Rock Flowage.*—Rock flowage is that portion of non-elastic deformation which occurs without fracturing. It includes rearrangement of minerals (recrystallization), formation of flow cleavage, migration of materials from limb to crest in a fold, elongation of components like ooids, pebbles,

fossils or amygdules. It is rarely entirely free from minute fracturing as seen in quartz grains, undulatory extinction or mineral cleavage. By flowage, the author means to describe the process which deforms beds without destruction of continuity of bedding. Intensity of flowage varies greatly and bedding may seem entirely unaffected or in intense deformation may be completely obliterated as, for instance, in crystalline schists. But inasmuch as no geologist places much confidence in measurements of stratigraphic thicknesses in recrystallized rocks, they are not here discussed. All formations included here are known as generally non-metamorphic and fossiliferous.

Flowage manifests itself in rearrangements of minerals at varying degrees of intensity, and in close relation to the folds in which it takes place. It culminates in flow cleavage which becomes prominent if the re-orientation is sufficiently intense. Systematic investigation of oolitic limestones west of South Mountain in Maryland has furnished quantitative data on the amount of deformation necessary to make cleavage visible. The minimum ratio of the long axis of an elongated ooid over the shortest axis is 1.25:1, if cleavage is barely visible in the field. A ratio of 2:1 shows excellent cleavage and it reaches 7:1 beyond which ooids cannot be measured because of their destruction.

In quartzites deformation is also visible and strongly elongated quartz grains have been measured. The micas, however, "lubricate" the system and protect the quartz. If deformation is sufficient, quartz schists occur.

A systematic investigation of flowage in sandstones and quartzites in the central Appalachians by R. Fellows is well under way. In the Weverton quartzite, elongations of well-rounded quartz grains reach 2:1 without recrystallization of the entire rock. Cleavage is coarse but well visible, bedding is uninterrupted.

In shales cleavage is readily formed and measurements of crinoid stems in the Martinsburg shale show elongations of 2:1 and more.

In several hundred thin sections studied so far, cleavage is without exception in the direction of flowage. It contains the largest and mean axis of the deformation and is perpendicular to the smallest axis. This relation can easily be seen in deformed ooids and can be measured accurately wherever known shapes are available. Its intensity increases with increasing axial ratio.

If this ratio is 4:1, the extension of an originally spherical body is 100% of its original diameter. The present grain is twice as long and one-half as wide. Other ratios permit determination of the proportional amount of deformation. This distortion takes place by flowage, and the individual units, their boundaries and primary structures can still be studied.

*Flowage, Cleavage and Bedding.*—Since flowage results in cleavage, such flow cleavage becomes an important criterion in the analysis of the dis-



tortion of a bed. Where cleavage transects bedding at high angles, the beds have thickened, as is well known in all anticlines or synclines. Crests of folds are usually thicker than the limbs and in these crests cleavage is at high angles to bedding. On the other hand, limbs of folds are thinner and cleavage is at low angles or parallel to the bedding planes. Competent beds thicken less than incompetent ones, show cleavage less distinctly, and deformation ratios are smaller. Measurements of oolites along the western foothills of South Mountain in Maryland and Pennsylvania show, as a rule, that cleavage transects bedding at high angles. In this region, which comprises the entire Cambro-Ordovician sequence, exaggerated thicknesses can be expected.

*Distorted Thicknesses.*—If a type section has been measured at that area of an anticline at which beds are transected by cleavage at large angles (greater than 45 degrees), it can safely be assumed to be exaggerated. Thicknesses measured along limbs of folds are most likely too thin and certain incompetent formations may even be missing if they are adjacent to competent ones. A comparison of such thicknesses has led to the conclusion that depressions and swells existed within the geosyncline which has received the sediments.

A few examples taken from the literature will show the reliability of such conclusions and the necessity of a careful structural analysis in each section which is measured.

The lower Cambrian Weverton sandstone or quartzite was so named by Keith<sup>1</sup> in 1894. This author estimated a thickness of 500 feet at Blue Ridge and 300 feet at Catoclin Mountain. Stose<sup>2</sup> in 1909 writes: "The thickness cannot be accurately determined but their relative positions are shown in the columnar sections. The total thickness of the formation, computed from dips and width of outcrop is about 1250 feet. The rocks are so sheared and metamorphosed that the original bedding and even character of the original sediment cannot be determined."

South Mountain is a highly asymmetric anticline overturned westward and the distortion of bedding mentioned by Stose points toward a high angle of intersection between bedding and cleavage. It seems far from coincidental that the estimates of the two authors vary between 1250 and 300 feet from the crest and limb of the same anticline within one formation.

The Harpers shale is above the Weverton sandstone and was named by Keith from Harpers Ferry, West Virginia. In the above-mentioned report that author states that "... in no one (of the exposures) can the thickness be measured with any degree of accuracy, for they are folded and twisted beyond description. At Harpers Ferry, where the lithological exhibition is complete, the section is a hopeless tangle. The cleavage planes dip 60 to 80 degrees to the southeast, but the bedding can readily be traced in

every direction and at every angle. No measure of thickness of any value whatever can be obtained here.

"Making a liberal allowance for unrepresented portions and judging somewhat from the breadth of its outcrops, a probable thickness of 1200 feet can be assigned to the formation."

Stose describes the Harpers formation on the road from Waynesboro to Monterey (Pennsylvania) within the Chambersburg quadrangle as follows: "In the Chambersburg quadrangle the most accurate determination of the thickness is obtained at the ends of the plunging anticlines northeast of Waynesboro and north of Fayetteville where the dips are low and uniform." The estimate given here is 2750 feet and the author added that "... in most places the rock is a schist."

The difference between the two estimates is more than 100% and in the new road cuts east of Waynesboro the cleavage intersects bedding at an angle of about 90 degrees.

In Adams County, Pennsylvania, Stose describes a similar situation: "... sericite ... is arranged parallel to the planes along which movement of the particles took place during compression and gives the rock its schistose character." Here cleavage intersects bedding at an angle of 90 degrees. The thickness of the formation is estimated as 3000 feet which is still more than near Waynesboro.

The Antietam schist overlies Harpers schist and Keith (14th report) states that the best locality for thickness measurements is at the north end of Blue Ridge. His estimate is between 500 and 700 feet at Front Royal. He states, however, that "... no better measures than those are known, and it would be hazardous to draw deductions from them." Stose comes to similar conclusions and estimates about 500 feet for the southern part and 800 feet in the northern part of the Chambersburg quadrangle. At White Rock the cleavage dips 35 degrees east, bedding 80 degrees west. The angle of 65 degrees points toward an exaggerated thickness.

The Tomstown dolomite was named after Tomstown north of Waynesboro by Stose. "The thickness computed from width of its outcrop and dip of its beds is about 1000 feet. The dip of the beds is 10 degrees south under red shale of the overlying Waynesboro formation." This thickness is also computed in an anticlinal nose. An investigation of oolite deformation at this locality shows intense elongation within the axial plane. The ratio is approximately 4:1, thus indicating an exaggerated thickness of 100%.

The Waynesboro formation as named by Stose overlies the Tomstown dolomite at Waynesboro. The beds are flat lying and were computed to be 1000 to 1750 feet thick in this area which is also a portion of the same anticline.

The Elbrook formation was named by Stose after the village of Elbrook, 5 miles northwest of Waynesboro (Chambersburg quadrangle). Here two computations from an asymmetrical syncline furnished a mean thickness of 3000 feet. Unfortunately, the author did not name the two values. This region is in an area in which oölite ratios vary between 1.5:1 and 2:1. Minor folding obscures the picture and an evaluation of this estimate is not yet possible.

The overlying Conococheague limestone was first measured in a section west of Scotland (Chambersburg quadrangle) as 1635± by Stose. Here bedding dips 70 to 80 degrees west and a strong transecting cleavage dips 20 to 30 degrees east. An excellent oölite bed is exposed east of the railroad bridge and its deformation was determined in thin sections. The average ratio of the longest over the shortest axes of ooids is 2.04:1 and the actual amount of extension in reference to a unit sphere is 55%. The bedding planes are intensely crumpled, offset and disarranged but still excellently visible. Inasmuch as the angle between maximum elongation and bedding is between 80 and 90 degrees, it stands to reason that the beds have been thickened by 50% and thus the original thickness exaggerated by at least 540 feet.

Many more examples can be cited for the whole paleozoic section of the Appalachian area. It seems strange that so many type sections have been computed or measured without structural considerations or description of the attitude of cleavage in relation to bedding. There seems little doubt that many such sections should be reexamined. Conclusions drawn from them are of rather doubtful value.

*Distorted Bedding.*—If cleavage as a plane in which flowage takes place transects bedding, it tends to distort it and finally obliterate it. Where cleavage is intense it provides avenues for the introduction of new materials like quartz, calcite or dolomite. The whole rock tends to appear different from the same formation elsewhere. Fossils become distorted or obliterated. Even changes in "lithology" may be produced by transecting cleavage planes. One of the best examples may be found in Hagerstown valley west of South Mountain uplift. The limestones west of the Martinsburg shale belt differ from those in the east probably less because of lithologic changes but because the deformational ratio and thus the intensity of cleavage increases from west to east. West of the Martinsburg shale it does not exceed 1.5:1 and to the east it reaches 7:1 in the Conococheague limestone. Many workers have also noted that in the east fossils are scarce!

*Conclusions.*—Wherever thicknesses are measured or computed, they are cited as observed in the field and listed. Thus after insertion into the literature, these values are used by others as basis for far-reaching conclusions, correlations and speculations. Once "established" such values

become factual data in spite of the original author's careful restriction or repeated warnings as, for instance, those cited above.

A study of the literature shows, however, that references to flow cleavage are extremely rare and that almost no description considers the distortion of beds due to folding and rock flowage as indicated by flow cleavage. It would be highly desirable to determine distortion or flowage quantitatively but this is unfortunately rarely possible. It is feasible where oölites, pebbles, fossils or other known units are present and distorted. But even where quantitative analysis is impossible, it seems essential to know the direction, intensity, quality and general behavior of flow cleavage in a measured or computed section.

About 12,000 feet of sediments have been measured or computed in the above-cited example. Cleavage transects bedding at high angles wherever these determinations were made. Oölite measurements in the limestones show elongations of 100% in this area and quartz grain elongation is 50%. It seems quite possible that the total thickness is, therefore, not 12,000, but 8000 feet or less.

Where sections are described in the literature without complete structural data, they represent present thicknesses but they do not necessarily represent thicknesses of deposition; they tell nothing about the original depth of the geosyncline, trough or furrow, or the shape of its floor; they do not permit calculation of rate of deposition; and in general do not permit deductions of any kind which assume that present thickness equals or resembles the thickness of that formation prior to folding.

<sup>1</sup> A. Keith, U. S. G. S., 14th Annual Report, p. 328, 1894.

<sup>2</sup> G. Stose, U. S. G. S., Chambersburg folio, 1909.

---

## *THE PERMUTATION GROUPS OF A GENERAL DEGREE*

BY G. A. MILLER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS

Communicated September 4, 1942

Although group theory as an autonomous science originated with the study of permutation groups these groups present many difficulties which have not yet been overcome notwithstanding the fact that when only groups of small degrees are considered the study of the permutation groups is remarkably simple. One reason why the permutation groups of small degrees were the first to receive much attention is that these groups are the only ones which require consideration in the theory of the algebraic equations of small degrees. As these equations naturally were the first to

be deeply studied by the early mathematicians it resulted that the properties of the permutation groups of small degrees attracted the attention of those working along this line long before a general theory of permutation groups received much attention.

In the case of the intransitive groups of the general degree  $n$  those in which each of the systems of intransitivity is of degree 2 are the simplest since they are obviously abelian and of type  $1^m$ . Hence their structure is relatively simple from the standpoint of abstract groups. In particular, the total number of such abstract groups whose orders do not exceed  $2^m$  is known to be  $m + 1$  if the identity is included among these possible groups for a particular value of  $m$ . On the other hand, the determination of all these permutation groups involves difficulties which increase rapidly with the increase of the value of  $n$  even in this relatively simple category of groups. It should be noted that while two abstract groups are regarded as the same group whenever they are simply isomorphic two permutation groups are regarded as the same only if they satisfy the additional condition that they can be transformed into each other by means of permutations on the letters involved.

Among the permutation groups which involve only transitive constituents of degree 2 those in which all the permutations besides the identity are of the same degree constitute a fundamental special category to which we shall at first restrict our attention. It should be observed that both the degree and the order of such a group are necessarily even numbers. In fact, the order is clearly always a power of 2 and hence it may be assumed to be of the form  $2^m$ . There is evidently one and only one such group of order 2 for every even value of  $n$  and hence it will be convenient to assume in what follows, unless the contrary is explicitly stated, that  $m > 1$ . It will be found, in particular, that 2 is the only order for which there exists such a group for every even degree.

Since the average number of letters in all the permutations of a group of degree  $n$  is always equal to this degree diminished by the number of the transitive constituents of the group it results that when each of the transitive constituents of the group of degree  $n$  is of degree 2 then the average number of letters in all of its permutations is  $n/2$ . It results, therefore, that when each of the permutations besides the identity is of the same degree and the order of the group is  $2^m$  the degree of the group must be a multiple of  $2^m - 1$  since each of the permutations besides the identity must contribute the same number of letters towards bringing the number of letters of the identity up to the average. As this average is  $n/2$  there results the following theorem: *In every permutation group in which all the permutations besides the identity are of the same degree this degree is an even multiple of  $2^m - 1$ , where  $2^m$  is the order of the group.*

In order to simplify the considerations which relate to the general case

as regards the number of the possible groups we shall at first restrict our attention to the simplest remaining case, namely, the case when the order of the group is 4. In this case  $2^m - 1$  is 3, and, according to the theorem noted in the preceding paragraph, the degree of the group is an even multiple of 3 and hence a multiple of 6. We may therefore represent it by  $6k$ . When  $k = 1$  there is one and only one such group, viz., the group of degree 6 and of order 4 which involves three permutations of degree 4. By establishing a 1,1 correspondence between two such groups we obtain a group of degree 12 in which all the permutations besides the identity are of degree 8. In a similar manner we can obtain a group of degree 18 in which each of the permutations besides the identity is of degree 12. This process may be repeated indefinitely so as to obtain a group corresponding to an arbitrary multiple of  $k$ . The fact that there is only one group in each case can perhaps be best seen by considering the general situation.

In general, the degree of the groups is  $2k(2^m - 1) = n$  and the average number of its letters is  $k(2^m - 1)$ . This is therefore the number of letters which its  $2^m - 1$  permutations which differ from the identity must make up for the identity to bring the number of its letters up to the average. That is, each of these permutations must make up  $k$  letters for the identity and its degree must therefore be  $k \cdot 2^m$ . The total number of letters in all the permutations of the group is  $k2^m(2^m - 1)$ . This proves the following theorem: *If each of the transitive constituents of a permutation group of degree  $n$  is of degree 2 the group is abelian and of order  $2^m$ . If, moreover, all the permutations of this group, besides the identity, are of the same degree then this degree is  $k2^m$ , and the degree of the group is  $2k(2^m - 1)$ . The total number of letters in all the permutations of the group is  $k2^m(2^m - 1)$ .*

To see that for every pair of integral positive values of  $k$  and  $m$  there is one and only such group it may be desirable to form the group in the following manner. Construct a permutation of degree  $k \cdot 2^m$  on  $k$  sets of  $2^m$  letters each. Then add thereto a similar permutation on one-half of each of these  $k$  sets of letters and the same number of additional letters. These two permutations generate a group of order 4. If  $m$  exceeds 2 divide each of the given half sets into two equal parts again, including the added half sets, and add an additional part to the permutation thus obtained. Continue this process until the total number of letters obtained is  $2k(2^m - 1)$ , or until only 2 letters are added to letters of the extended sets. From this construction it results directly that there is only one such permutation group. That is, the following theorem results therefrom: *For every pair of positive integral values of  $k$  and  $m$  there is one and only one permutation group of order  $2^m$  and of degree  $2k(2^m - 1)$  which has the property that each of its transitive constituents is of degree 2 and that all of its permutations besides the identity are of the same degree.*

This interesting category of permutation groups has a number of striking



properties due to its unique character. In particular, if two subgroups of such a group are of different orders they are also of different degrees and the subgroup of the larger order is also of the larger degree. This results directly from the method of constructing these subgroups as noted in the preceding paragraph. From the same method it follows that when two subgroups are of the same order they are also of the same degree. When the order of such a group exceeds 2 the group is positive since at least half of the permutations of every permutation group are positive. If the order of a proper subgroup of such a group is  $2^{m_1}$ ,  $m > m_1$ , then it follows that the degree of this subgroup is an even multiple of  $2^{m_1} - 1$ . For instance, when  $m = 3$  and  $k = 1$  the degree of the group is 14 and the degree of the subgroups of order 4 is 12 while the degree of the subgroups of order 2 is 8.

It was noted above that for every even degree, there is at least one such group, viz., the group of order 2, and that this is also the only such group whose degree is a power of 2, while for degree 6 there are two such groups which are of orders 2 and 4, respectively. The lowest degree for which there are more than two such groups is 42. In this case there is a group of each of the orders 2, 4, 8. There is clearly no upper limit to the number of such groups which may exist for a given even degree and involve exactly this number of letters since this number is equal to the number of its different factors of the form  $2^m - 1$  for different values of  $m$ . While this category of groups is not very dense the fact that there is no upper limit for the degrees of the groups which belong to it and the great simplicity of the determination of these groups as permutation groups make the category of theoretic interest. As their orders increase the difference in the degrees of these groups increases and the category emphasizes that only a slight penetration into the general theory of permutation groups has yet been made notwithstanding the age of this theory.

---

## THE STRUCTURE OF ALGEBRAIC MODULS

BY GERALD HARRISON

CALIFORNIA INSTITUTE OF TECHNOLOGY

Communicated August 4, 1942

*Introduction.*—The abstract notion of a lattice, together with added structural restrictions such as the modular identity, first appeared in Dedekind's study of algebraic numbers. Dedekind showed that the free modular lattice generated by three elements is of order twenty-eight,<sup>1</sup> and exhibited sets of three quadratic moduls which actually generate it.

Though the free modular lattice on four generators is infinite, the free lattice generated by any finite number of submodules of a modul of finite order is finite. The finite basis together with a certain simplicity in the covering relation leads to a direct product decomposition for such moduls into lattices of a comparatively simple type, but the distinctive property of this decomposition is its multiplicative character.

Algebraic moduls have a naturally definable multiplication which leads, as Dedekind showed, to the ideal theory of algebraic number rings, and suggests a further study of the multiplicative properties of moduls. Although algebraic moduls satisfy certain lattice-multiplicative identities (e.g.,  $a(b \cup c) = ab \cup ac$ ), the connection between the lattice property of inclusion and the multiplication is much looser than in ideal theory.<sup>2</sup>

*Notation and Definitions.*—We use the usual  $\cup$  and  $\cap$  notation of lattice theory.<sup>3</sup>  $a$  covers  $b$  means  $a \neq b$  and that  $a \supseteq c \supseteq b$  implies  $a = c$  or  $c = b$ . The modul with basis elements  $\omega_1, \omega_2, \dots, \omega_n$ , all of which are of infinite order, is denoted by  $(\omega_1, \omega_2, \dots, \omega_n)$ . It is known that if  $\omega_1, \omega_2, \dots, \omega_n$  are linearly independent over the integral domain of rational integers, the submoduls of  $(\omega_1, \omega_2, \dots, \omega_n)$  can be written in the canonical form

$$(a_{11}\omega_1, a_{21}\omega_1 + a_{22}\omega_2, \dots, a_{n1}\omega_1 + a_{n2}\omega_2 + \dots + a_{nn}\omega_n), \quad (1)$$

where the  $a_{ij}$  are rational integers, and  $a_{jk} \leq a_{ik}$  if  $j < i \leq n$ ,  $k = 1, 2, \dots, n-1$ , the equality occurring only if  $a_{jk} = a_{ik} = 0$ . The norm of this modul is  $a_{11}a_{22} \dots a_{nn}$ . If  $p$  is any rational integral prime, the modul (1) covers the following moduls:

$$\begin{aligned} & (pa_{11}\omega_1, (a_{21} + k_{n-1, n-1}a_{11})\omega_1 + a_{22}\omega_2, \dots, (a_{n1} + k_{n-1, n-1}a_{11})\omega_1 + a_{n2}\omega_2 + \\ & \quad \dots + a_{nn}\omega_n) \\ & (a_{11}\omega_1, pa_{21}\omega_1 + pa_{22}\omega_2, (a_{31} + k_{n-2, n-2}a_{21})\omega_1 + (a_{32} + k_{n-2, n-2}a_{22})\omega_2 + a_{33}\omega_3, \dots, \\ & (a_{n1} + k_{n-2, n-2}a_{21})\omega_1 + (a_{n2} + k_{n-2, n-2}a_{22})\omega_2 + a_{n3}\omega_3 + \dots + a_{nn}\omega_n) \\ & \dots \dots \dots \\ & (a_{11}\omega_1, a_{21}\omega_1 + a_{22}\omega_2, \dots, pa_{n-1, n-1}\omega_1 + pa_{n-1, n-1}\omega_2 + \dots + pa_{n-1, n-1}\omega_n, \\ & (a_{n1} + k_{11}a_{n-1, n-1})\omega_1 + (a_{n2} + k_{11}a_{n-1, n-1})\omega_2 + \dots + (a_{nn} + k_{11}a_{n-1, n-1})\omega_n) \\ & (a_{11}\omega_1, a_{21}\omega_1 + a_{22}\omega_2, \dots, pa_{n1}\omega_1 + pa_{n2}\omega_2 + \dots + pa_{nn}\omega_n), \end{aligned}$$

where  $k_{ij}$ ,  $j \leq i \leq n-1$  take on any integer values between and including 0 and  $p-1$ . We say the modul (1) covers these moduls with respect to  $p$ . Thus each modul covers  $(p^n - 1)/(p - 1)$  moduls with respect to each rational integral prime  $p$ , and these are those and only those moduls whose norm is greater by the factor  $p$ . For example, in the case  $n = 2$  the modul  $(a_{11}\omega_1, a_{21}\omega_1 + a_{22}\omega_2)$  covers the moduls

$$(pa_{11}\omega_1, (a_{21} + k_{11}a_{11})\omega_1 + a_{22}\omega_2), \quad k_{11} = 0, 1, \dots, p-1$$



and

$$(a_{11}\omega_1, pa_{21}\omega_1 + pa_{22}\omega_2)$$

with respect to  $p$ . As  $p$  runs through all the primes we obtain just once each of the moduls covered by a given modul. If we start with the unit modul  $(\omega_1, \omega_2, \dots, \omega_n)$  and form all possible successive coverings with respect to a fixed prime  $p$ , we obtain a sublattice of the lattice of all submoduls of  $(\omega_1, \omega_2, \dots, \omega_n)$  whose elements are the submoduls of  $(\omega_1, \omega_2, \dots, \omega_n)$  whose norm is a power of  $p$ . We call such a lattice a  $p$ -lattice. The "top" of the Hasse diagram of a  $p$ -lattice is easy to construct in the quadratic case.

*Decomposition Theorems.*—THEOREM I. Let  $p_1, p_2, \dots, p_i, \dots$  be the set of all rational integral primes, corresponding to each of which is a  $p_i$ -lattice of  $(\omega_1, \omega_2, \dots, \omega_n)$ . Then if  $m$  is an arbitrary submodul of  $(\omega_1, \omega_2, \dots, \omega_n)$ ,

$$m = m_{p_1} \cap m_{p_2} \cap \dots \cap m_{p_i} \cap \dots \quad (2)$$

where  $m_{p_i}$  is the unique modul of the  $p_i$ -lattice which is the meet of all moduls of the  $p_i$ -lattice which contain  $m$ . This representation is unique, and if

$$n = n_{p_1} \cap n_{p_2} \cap \dots \cap n_{p_i} \cap \dots$$

is a decomposition of  $n$ , then

$$\begin{aligned} m \cap n &= (m_{p_1} \cap n_{p_1}) \cap (m_{p_2} \cap n_{p_2}) \cap \dots \cap (m_{p_i} \cap n_{p_i}) \cap \dots \\ m \cup n &= (m_{p_1} \cup n_{p_1}) \cap (m_{p_2} \cup n_{p_2}) \cap \dots \cap (m_{p_i} \cup n_{p_i}) \cap \dots \end{aligned}$$

Thus the lattice of all submoduls of  $(\omega_1, \omega_2, \dots, \omega_n)$  is the direct product of the  $p_i$ -lattices.<sup>4</sup>

The next theorem has been proved for quadratic moduls, but probably holds in the  $n$ th degree case.

THEOREM II. Let the unit modul have the form  $(1, \omega)$  where  $\omega$  is a quadratic algebraic integer. If  $m$  and  $n$  are submoduls of  $(1, \omega)$  having the representations described in Theorem I, i.e.,

$$\begin{aligned} m &= m_{p_1} \cap m_{p_2} \cap \dots \cap m_{p_i} \cap \dots \\ n &= n_{p_1} \cap n_{p_2} \cap \dots \cap n_{p_i} \cap \dots \end{aligned}$$

then

$$mn = m_{p_1}n_{p_1} \cap m_{p_2}n_{p_2} \cap \dots \cap m_{p_i}n_{p_i} \cap \dots$$

*Conclusions.*—Some corollaries of Theorems I and II are:

1. If two moduls are such that at least one of the components with respect to each prime is the unit modul, then their product is an ideal. For example, the product of  $(3, 2 + a\omega)$  and  $(7, 2 + \omega)$  is an ideal in any quadratic ring.

2. A submodul of  $(1, \omega)$  is a ring or ideal if and only if each of its com-

ponents is a ring or ideal, respectively. This reduces the study of the ideal theory of  $(1, \omega)$  to the ideal theory of the  $p_r$ -lattices.

3. The further study of the multiplicative and lattice theoretic decompositions of moduls may be restricted to their  $p_r$ -lattice components.

A more detailed account of the results of this note will be published elsewhere. I wish to express my sincere thanks to Professor Morgan Ward for his advice and encouragement in the above work.

<sup>1</sup> *Gesammelte mathematische Werke, Zweiter Band, Über die von drei Moduln erzeugte Dualgruppe*, pp. 236-271.

<sup>2</sup> For example, the product of two moduls may contain, be contained by, or not be related through containing to its factors. There exist moduls  $a$  such that  $a$  does not contain  $a^s - 1$  but  $a \supset a^s$ ,  $s$  some integer  $> 1$ , or even  $a$  does not contain  $a^s - 1$ ,  $a = a^s$ , in which case  $a$  generates a cyclic group of order  $s - 1$  under multiplication.

<sup>3</sup> Birkhoff, G., *Lattice Theory*, 1940.

<sup>4</sup> Every representation<sup>3</sup> has only a finite number of non-unit components.

## SYNTHETIC SOLUTION OF THE INVERSE PROBLEM OF DYNAMICS

BY EDWARD KASNER AND JOHN DE CICCIO

DEPARTMENTS OF MATHEMATICS, COLUMBIA UNIVERSITY AND ILLINOIS INSTITUTE OF TECHNOLOGY

Communicated August 18, 1942

1. *Introduction.*—In a given plane positional field of force, there are  $\infty^3$  dynamical trajectories. The direct problem of dynamics is to find the trajectories when the field is given. A completely characteristic set of five properties of any such system has been given in the Princeton Colloquium. The inverse problem proposed and solved analytically by Kasner is to find the field of force when either all or only a certain subset of its trajectories are known. The result is that *a field of force is, in general, completely determined except for a constant factor by four simple ( $4 \infty^1$ ) families of trajectories.* The way in which this result is established may be described as a method for *the geometric exploration of a field of force.*<sup>1</sup>

In this paper, we propose to give a purely synthetic construction of the essential problem<sup>2</sup> of the preceding theorem, which is to find the direction of the force acting at a point  $O$  when there are given four trajectories with distinct tangents at  $O$ . This is accomplished by the introduction of a new series of lineal elements, which we call a *limaçon series* (of the second kind). This is different from the limaçon series (of the first kind) discussed elsewhere in a different connection.<sup>3</sup> The direction of the field of force is uniquely determined if the four associated inverted focal elements (relative

to an arbitrarily chosen direction) of the four given trajectories do not lie on a limaçon series. If they do lie on a limaçon, the problem is indeterminate.

2. *The First Three Properties of a Complete System of  $\infty^1$  Dynamical Trajectories.*—In order to solve the problem stated above, it is necessary to state the first three of the characteristic set of five properties of a complete system of dynamical trajectories.

*Property I:* If for each of the  $\infty^1$  trajectories passing through a given point in a given direction, we construct the osculating parabola at the given point, the locus of the foci of these parabolas is a circle passing through that point.

*Property II:* The focal circle that corresponds, according to Property I, to a lineal element  $E$ , is so situated that  $E$  bisects the angle between the tangent to the circle and a certain direction fixed for the given point. (This is the direction of the force acting at the given point.)

*Property III:* The envelope of the  $\infty^1$  focal circles is a circle (in a general position).

3. *The Geometric Determination of the Direction of the Force by Four Trajectories with Distinct Tangents.*—Now we proceed to examine our inverse problem. Let four trajectories with distinct tangents at the point  $O$  be given. Consider an arbitrary direction at  $O$ , and let us see if it can be the direction of the force acting at that point. Take the image of this direction in the tangent to the first of the given curves. By Properties I and II, the focal circle must be the one through  $O$  in the direction so obtained and through the focus of the corresponding osculating parabola. Doing this for each of the four curves, we obtain four focal circles. From Property III, we find that this test is correct if there exists a circle touching these four. We have then a purely geometric problem: to find a direction at  $O$  such that the four circles constructed by means of it shall admit a common tangent circle.

We may simplify this problem by inverting the configuration considered with respect to  $O$ . Then there are, instead of four focal circles, four straight lines which are to be concyclic. As the direction tested is changed, these rotate simultaneously through equal angles about four fixed points, namely, those obtained by inverting the four foci.

Take an arbitrary oriented direction for trial. Construct for each of the four inverse foci, a direction parallel to the tangent through the point  $O$  of the corresponding focal circle. We thus obtain four oriented lineal elements  $E_1, E_2, E_3, E_4$ , one at each of the inverse foci. The problem<sup>1</sup> is to turn these elements (each about its own point) through some angle  $\alpha$ , so that the new elements  $E_1', E_2', E_3', E_4'$  shall have concyclic lines. Before discussing the synthetic construction of this problem, it is necessary to consider our new limaçon series.

4. *The Limaçon Series.*—This new series is given geometrically as follows. Let a circle  $\Gamma$  be denoted by its center  $C$  and a fixed point  $\pi$  on its circumference. Draw any (oriented) line  $l$  through  $\pi$  intersecting this circle  $\Gamma$  again in  $Q$ . Let  $E_Q$  be the lineal element with point at  $Q$  and direction that of  $l$ . Upon applying the slide  $S_\delta$ , and then the turn  $T_\phi$  to  $E_Q$ ,<sup>4</sup> we obtain the lineal element  $E$ . The set of  $\infty^1$  lineal elements  $E$  form our limaçon series  $L$ .

The point  $\sigma$  on the circumference of the circle  $\Gamma$  so constricted that the arc  $\pi\sigma$  has the angular measure  $2\phi$  is the center of the circle to which all the lines of the lineal elements of the limaçon series  $L$  are tangent. The radius of this circle is  $\delta \sin \phi$ . Thus the line-union of  $L$  is a circle, and the point-union is a limaçon.

For the limaçon series  $L$ , we shall call  $\Gamma$  the *base circle*, the point  $\pi$  the *pole*, the point  $\sigma$  the *line-center*,  $\delta$  the *determining length*, and  $\phi$  the *determining angle*. There are  $\infty^6$  limaçon series in the plane.

Three elements  $E_1, E_2, E_3$ , which are not all parallel, determine a unique limaçon series.

The pole  $\pi$  of our limaçon series  $L$  is determined as follows. Construct the circle  $C_k$  containing the points of  $E_i$  and  $E_j$  and the point of intersection of the lines of  $E_i$  and  $E_j$ . The three circles so obtained intersect in a single point, which is the pole  $\pi$  of  $L$ .

The base circle  $\Gamma$  is obtained in the following way. Let  $l_i$  be the line through the point of  $E_i$  perpendicular to the line joining  $\pi$  to the point of  $E_i$ . Let  $\pi'$  denote the center and  $\delta$  the radius of the unique circle tangent to  $l_1, l_2, l_3$ . The base circle of our limaçon series  $L$  possesses the line segment  $\pi\pi'$  as diameter.

Of course, the radius  $\delta$  is the determining length; and  $\phi$ , the constant angle that  $E_i$  makes with the line joining  $\pi$  to the point of  $E_i$ , is the determining angle of our limaçon series.

The center  $\lambda$  of the turbine containing any two elements  $E_1$  and  $E_2$  of a limaçon series  $L$  is on the base circle of  $L$ . Also the circle determined by the points of  $E_1$  and  $E_2$ , and the point of intersection of their lines, contains the pole  $\pi$  and the center  $\lambda$  of the turbine determined by  $E_1$  and  $E_2$ .

Any two limaçon series with congruent base circles are equivalent under the whirl-motion group  $G_6$ .

5. *The Problem of Making the Lines of Four Elements Concylic by the Application of a Turn  $T_\alpha$ .*—Now we shall consider the problem outlined at the end of Section 3. Let  $E_1, E_2, E_3, E_4$  be four elements not all on one limaçon series. Let  $L_3$  and  $L_4$  be the limaçon series determined by the elements  $(E_1, E_2, E_3)$  and  $(E_1, E_2, E_4)$ , respectively. The base circles of  $L_3$  and  $L_4$  intersect in the point  $\lambda$ , the center of the turbine determined by  $E_1$  and  $E_2$ , and another point  $X$ . The angle of the turn  $T_\alpha$  is given by one-half the arc from the line-center  $\sigma_3$  to the point  $X$ , that is, arc  $\sigma_3 X$  on the

base circle of  $L_3$ . The application of this turn  $T_a$  will carry our two limaçon series  $L_3$  and  $L_4$  into two new ones with same base circles and same poles but with a common line-union which is a circle with center at  $X$ .

The application of the inverse turn  $T_{-a}$  to the trial direction, given in the last paragraph of Section 3, will yield the direction of the force.

Of course, if the four elements are on one limaçon series, the direction of the force is indeterminate.

6. *Conclusion.*—We wish to summarize our results in the purely geometric form. For this purpose, it is found appropriate by Properties I and II, to introduce *the associated focal element relative to a fixed direction  $F$*  of a differential element of third order through a given point  $O$ . At  $O$ , let  $G$  be the lineal element which is the symmetrical image of the fixed lineal element  $F$  through the direction of the third order differential element. The lineal element  $f'$  through the focus of the osculating parabola of the third order element so that  $f'$  and  $G$  are cocircular is called the associated focal element relative to the fixed direction  $F$ . Inverting  $f'$  with respect to  $O$ , we obtain the *associated inverted focal element  $f$*  relative to a fixed direction  $F$  of the given third order element.

The direction of the force is uniquely determined by four differential elements of the third order with distinct tangents at  $O$  if the associated inverted focal elements relative to a fixed direction  $F$  do not all lie on one limaçon series. Otherwise the direction is indeterminate.

If we do not apply an inversion, then the direction of the force at  $O$  is uniquely determined if and only if the four associated focal elements relative to a fixed direction  $F$  do not lie on an *extended* limaçon series. This new series may be defined as follows. Through the point  $O$  and the node of a limaçon, draw a circle intersecting the limaçon in a point  $P$ . (There are two such points  $P$ .) Construct the lineal element  $E'$  of this circle with point at  $P$ . The application of the turn  $T_\phi$  to this element  $E'$  yields the element  $E$ . The set of all such elements  $E$  is termed an extended limaçon series.

Thus we have solved the inverse problem of dynamics in a purely geometric manner by use of the concept limaçon series together with concept turbines. All the steps can be carried out with straight edge and compass. Hence if we are given a photograph of the entire system of trajectorial curves (or a sufficiently large subset of the system) generated by some unknown positional field of force, without any record of motion or time, we can actually find the law of the field of force. On the other hand, a photograph of the lines of force would not be sufficient to determine the field.

<sup>1</sup> Kasner, "Differential Geometric Aspects of Dynamics," *Princeton Colloquium Lectures*. Published by the Am. Math. Soc. (1913, 1934). Also *Trans. Am. Math. Soc.*, 1906-1910. Geometric solutions of the problem of four lineal elements have been indicated orally to the senior author by J. Wedderburn and J. Douglas. Our dynamical

problem should also have a purely projective solution, since the five fundamental properties of trajectories are essentially projective. See A. Terracini, *Rivista di Matematica*, 2, 245-329 (1941), for a new elegant discussion of the properties.

<sup>2</sup> If we know the path of a particle and also the direction of the force acting at each of its points, then, assuming the magnitude arbitrarily at one point, it is completely determined at all points. This is a quadrature problem.

<sup>3</sup> De Cicco, "The Geometry of Fields of Lineal Elements," *Trans. Am. Math. Soc.*, 47, 207-229 (1940).

<sup>4</sup> For the definitions of turns and slides, see Kasner, "The Group of Turns and Slides and the Geometry of Turbines," *Am. Jour. Math.*, 33, 193-202 (1911). Also Kasner and De Cicco, "Geometry of Turbines, Flat Fields, and Differential Equations," *Ibid.*, 59, 545-563 (July, 1937).

## ON THE THEORY OF ANALYTIC CURVES

BY HERMANN AND JOACHIM WEYL

INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.

Communicated August 11, 1942

In a paper of great importance and beauty,<sup>1</sup> L. V. Ahlfors has simplified and vastly expanded the theory of meromorphic curves inaugurated by the authors of this note.<sup>2</sup> In the meantime the younger of us had generalized the first and second main theorems of our theory to arbitrary *analytic curves* whose parameter  $p$  varies over a given Riemann surface  $\mathfrak{M}$ .<sup>3</sup> The following observations will show how by a proper adaptation of Ahlfors' method this more general theory may be brought up to the same degree of completion as the special case of the meromorphic curves.

As in reference 3 let  $G$  denote any compact region of  $\mathfrak{M}$  which surrounds a given nucleus  $G_0$ , and  $\bar{G}$  its complement. We speak of the *condenser*  $G$  whose outer conductor is  $\bar{G}$  and whose inner conductor  $G_0$  carries unit charge. Let  $\varphi = \varphi_G$  be the potential, which vanishes in  $\bar{G}$  and assumes a constant value  $R$  in  $G_0$ . To be precise, the space of the condenser is  $G^* - G_0^*$  where  $G_0^*$ ,  $G^*$  are defined by  $\varphi = R$ ,  $\varphi > 0$ , respectively; for the sake of simplicity we identify  $G_0^*$ ,  $G^*$  with  $G_0$ ,  $G$ . The constant  $R$  is called the potential (or the reciprocal capacity) of the condenser. Ahlfors' formula (19) for the order function  $T_i$  is generalized as follows. Let  $x(p)$  be our curve and  $t$  a local uniformizing parameter at any point of the Riemann surface. Form

$$X_t^i = [x, dx/dt, \dots, d^{i-1}x/dt^{i-1}],$$

$$S_t^i = 2 |X_t^{i-1}|^2 \cdot |X_t^{i+1}|^2 / |X_t^i|^4.$$

The integral element  $S_i dt d\bar{t}$  is invariant; if one uses any analytic differential  $dz$  instead of  $dt$  one gets an  $S_i$  connected with  $S_t$  by the equation

$$S_l |dz/dt|^2 = S_l.$$

We arrive at this expression for  $T_l$ ,

$$T_l = \int \int \varphi S_l^l \cdot dt \bar{dt}, \quad (1)$$

the integral extending over the entire Riemann surface (or over  $G$  only).

Our basic idea is to decompose the space of the condenser into thin layers by the equi-potential lines  $\varphi = \varphi_1$ . For a fixed value  $\varphi_1 = R - r$  the inequality  $\varphi_1 \leq \varphi \leq R$  defines a region  $G_r$ , and the condenser  $G_r$  has the potential  $R - \varphi_1 = r$ . The harmonic function  $\varphi$  is the real part of an analytic function  $\varphi - i\sigma = f$  in  $G - G_0$  which, to be sure, is not single-valued; but the differential  $df$  is, and this is all that counts. Use  $df$  instead of  $dt$  in that part of the integral (1) which extends over  $G - G_0$ . Omitting the index  $l$  we form the integral

$$Q(r) = \int S_r \cdot d\sigma \quad (d\sigma \geq 0)$$

along the line  $C_r: \varphi = R - r$ . (Here the differentiation  $d/df$  in  $S_r$  really amounts to  $d/d\sigma$ .) The flux through the line,  $\int d\sigma$ , is the total charge 1 (not  $2\pi$ ; being mathematicians, we use the Heaviside units). The formula (1) now reads

$$\begin{aligned} T &= A_0 R + \int_0^R \varphi \cdot Q(R - \varphi) \cdot d\varphi \\ &= A_0 R + \int_0^R (R - r) \cdot Q(r) \, dr, \end{aligned}$$

where  $A_0$  is the integral of  $S_l dt \bar{dt}$  over the nucleus  $G_0$  and hence independent of  $G$ . Applying the formula to the condenser  $G_r$  we obtain for  $T(G_r)$ :

$$T(r) = A_0 r + \int_0^r (r - \rho) Q(\rho) \, d\rho.$$

This proves at once that  $T(r)$  is a (positive increasing) *convex* function of the variable  $r$  and that

$$d^2 T/dr^2 = Q(r). \quad (2)$$

The second main theorem has been formulated in reference 3 in two different forms, first on the basis of a given "meromorphic" function  $z$  on  $\mathfrak{M}$ , and then by means of the intrinsic non-Euclidean metric on  $\mathfrak{M}$  to which the theory of uniformization leads. In the first form the meromorphic character of the differential  $dz$  in  $G$  is sufficient, and it is not necessary to use the same  $dz$  for each domain  $G$ . The really important relation arises from the choice  $dz = df$ . This differential is not defined inside  $G_0$ . Hence one must apply the fundamental formula to  $G - G_0$ , and the separate treatment of  $G_0$  brings in a topological moment. In the notation used loc. cit., we arrive at the *Plücker formulas for analytic curves*

$$V_l + (T_{l+1} - 2T_l + T_{l-1}) = \Omega_l(r) |_0^R + 2\pi\eta$$



in which the compensating term  $\Omega_l(r)$  is the integral

$$\Omega_l(r) = \frac{1}{2} \int \log S_f' \cdot d\sigma \quad (d\sigma \geq 0)$$

taken along the line  $C_r$  and

$$\eta = \sum_{\mathfrak{m}} \varphi(\mathfrak{m}) - \nu_0 R.$$

The last sum extends over the "critical points"  $\mathfrak{m}$ , i.e., the zeros of  $df$  inside  $G - G_0$  (or, what is the same, the zeros of the electric field strength  $-\text{grad } \varphi$ ).  $\nu_0$  is an integer, namely the Euler characteristic of  $G_0$ . It is clear that  $\eta$  depends on  $G$ , but neither on the index  $l$  nor on the curve  $x(p)$ . It corresponds to the term  $2p - 2$  in Plücker's formulas for an algebraic curve of genus  $p$ . The value of  $\eta$  is little influenced by whether or not one includes in the sum critical points  $\mathfrak{m}$  near the outer wall of the condenser where  $\varphi = 0$ . But it seems to violate the law of continuity at the inner wall. However, according to M. Morse's now classical relation,

$$\sum_{\mathfrak{m}} 1 = \nu_0 + \nu,$$

where  $-\nu$  is the Euler characteristic of  $G$ . Hence one can write

$$\eta = -\sum_{\mathfrak{m}} (R - \varphi(\mathfrak{m})) + \nu R,$$

a formula in which the critical points near the *inner* wall  $\varphi = R$  count very little. Notice the inequality

$$-\nu_0 R \leq \eta \leq \nu R.$$

Again dropping the index  $l$  we have

$$2\Omega(r) \leq \log Q(r).$$

The potential  $R(G)$  of the condenser  $G$  increases if  $G$  is enlarged, and hence converges either to  $\infty$  or to a finite limit  $R_0$  under exhaustion of  $\mathfrak{M}$  by  $G$  (case of infinite or finite total potential). Choose a number  $\kappa > 1$ . Considering that  $T'(r) \geq A_0$  and thus  $T(r) \geq A_0$  as soon as  $r \geq 1$  we find in familiar fashion from (2) that an inequality

$$\log Q(r) > \kappa^2 \cdot \log T(r) + (\kappa + 1) \log C$$

with a given positive constant  $C$  cannot hold throughout a subinterval of  $1 \leq r \leq R$  of length  $B/C$  where

$$2 \int_{A_0}^{\infty} y^{-\kappa} dy = 2A_0^{-(\kappa-1)} / (\kappa - 1) = B$$

is independent of  $G$ . Hence *in the case of infinite total potential we see that there will be values  $r$  in a boundary strip  $R(G) - \beta \leq r \leq R(G)$  of preassigned width  $\beta$  for which the inequality*

$$2\Omega(r) \leq \kappa^2 \cdot \log T(r) + (\kappa + 1) \log B/\beta$$



holds, as soon as  $R(G) \geq \beta + 1$ . With a given constant  $b > 1/2$  this fact prevents a relation like

$$\Omega(G) \geq b \cdot \log T(G)$$

from holding for all sufficiently large domains  $G$ ; but it says a good deal more about the behavior of  $\Omega(G)$  with respect to  $\log T(G)$ .

In the case of finite total potential  $R_0$  the function  $\varphi_G$  tends to a limit  $\varphi$  with the exhaustion of  $\mathfrak{B}$  by  $G$ , and it is then natural to use only the regions  $G_r (0 \leq r < R_0)$  defined by  $R_0 - r \leq \varphi \leq R_0$ . Almost everywhere, i.e., with the exception of an  $r$ -set over which the integral of  $(R_0 - r)^{-1}$  is finite, the inequality

$$2\Omega(r) < \kappa^2 \log T(r) + (\kappa + 1) \log (R_0 - r)^{-1}$$

will hold.

In establishing the defect relations by Ahlfors' procedure, let us stick to our use of the symbol  $\omega$ , as explained for the meromorphic case by formula (5.2) of reference 2; it differs from Ahlfors' usage by an arbitrary positive factor  $K$  which he adds to the function  $\theta$ . In handling arbitrary analytic curves one had better abstain from Ahlfors' trick of tying up the exponent  $\alpha < 1$  with the outer circle  $R_n$  of the annular condenser by the relation  $1 - \alpha = 1/T(R)$ . We include the cases of *non-general position* as in reference 3, section 7, by attaching non-negative weights  $\lambda(E^n)$  to the several  $h$ -spreads  $E^n$  of the given finite set. Ahlfors' method works best for the lowest case of a finite set of  $l$ -spreads or "points"  $c$ . Set

$$m_l(r; c) = \int \log \frac{|X^l| \cdot |c|}{|[X^l, c]|} \cdot d\sigma,$$

the integral running over the curve  $C_r$ . Suppose that the weights  $\lambda(c)$  attached to the points  $c$  of the given finite set will load no  $l$ -spread by more than 1. Then

$$\sum_c \lambda(c) \{ \alpha m_l(r; c) - m_{l-1}(r; c) \} + \Omega_l(r) = 1/2 \omega(T_l).$$

In particular for  $l = 1$  the first term of the left member may be written as

$$\alpha \cdot \sum_c m_1(r; c)$$

under the assumption that the points  $c$  are *distinct*. It is remarkable that no further restriction on the position of the points  $c$  is required. The cases  $h > 1$  in which  $[X^l, E^n]$  takes the place of  $[X^l, c]$  can also be treated with fixed exponents. Ahlfors' second set of defect relations arises from application to the dual curve of the first set thus obtained.

A more detailed account will probably be published in a planned monograph on analytic curves in the *Annals of Mathematics Studies*.

<sup>1</sup> Ahlfors, L. V., "The Theory of Meromorphic Curves," *Acta Societatis Scientiarum Fennicae*, Nova Series A, Tom. III, No. 4 (1941).

<sup>2</sup> Weyl, H. and J., *Ann. Math.*, **39**, 516-538 (1938).

<sup>3</sup> Weyl, J., *Ibid.*, **42**, 371-408 (1941).

## ON THE ANALYTICAL THEORY OF SEMI-GROUPS

BY EINAR HILLE

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY

Communicated September 2, 1942

1. Let  $E$  be a complex Banach space,  $\{T_s\}$  a one-parameter family of bounded linear transformations on  $E$  to  $E$  defined for  $s > 0$  and having the semi-group property

$$T_s T_t = T_t T_s = T_{s+t}, \quad s > 0, t > 0. \quad (1)$$

In an earlier note<sup>1</sup> the author announced representation theorems for such semi-groups. The present note elaborates one of these results and is further concerned with the behavior of  $T_h$  for small positive  $h$ .

2. We start with a result on the degree of approximation of  $I$  by  $T_h$ . We write  $A_h = (1/h)[T_h - I]$  and put  $Ax = \lim_{h \rightarrow 0} A_h x$  whenever the limit exists in the strong sense. The set of elements  $x$  for which the limit exists is a linear subspace  $D(A)$  which may reduce to the zero-element.

**THEOREM 1.** *Let  $\|T_s\|$  be bounded in every finite interval and put  $\sup_{0 < s < \omega} \|T_s\| = M(\omega)$ . If for a particular  $x$ ,  $\|(T_h - I)x\| \rightarrow 0$  in such a manner that  $\liminf_{h \rightarrow 0} \|A_h x\| = 0$  then  $Ax \equiv 0$ , i.e.,  $T_s x \equiv x$  for all  $s > 0$ . If  $\|(T_h - I)x\| \rightarrow 0$  with  $h$  for all  $x \in E$  and  $y \in D(A)$ , then  $\|A_h y\| \leq M(h)\|Ay\|$  or  $\|(T_h - I)y\| \leq hM(h)\|Ay\|$ .*

Roughly speaking the theorem asserts that only invariant elements  $x$  admit of an approximation by  $T_h x$  which is of a higher degree than the first in  $h$  and that first degree approximation is reached by all elements of  $D(A)$ . The boundedness assumption on the norm is satisfied if, for instance,  $\|T_s\|$  is measurable and  $\limsup_{h \rightarrow 0} \|T_h\| < \infty$ . If  $s = nh + \delta$ ,  $0 \leq \delta < h$ , a simple calculation gives

$$sA_h x = \delta A_\delta x + (T_\delta + T_{\delta+h} + \dots + T_{\delta+(n-1)h})hA_h x,$$

whence

$$\|A_s x\| \leq M(s) \liminf_{h \rightarrow 0} \|A_h x\| \quad (2)$$

from which the first assertion follows. A simple case of this part of the theorem was proved in an earlier paper.<sup>2</sup> For the second part we note that if  $y \in D(A)$  so does  $T_s y$  and  $AT_s y = T_s A y = \frac{d}{ds} T_s y$  is continuous on the right in  $s$ . Hence

$$\|(T_h - I)y\| = \left\| \int_0^h T_s A y ds \right\| \leq hM(h)\|Ay\|. \quad (3)$$

This theorem has a number of consequences in various branches of analysis, in particular in the theory of singular integrals and the summation of orthogonal series. Abel-Poisson summability of Fourier series provides an interesting example.

**THEOREM 2.** *If  $f(u) \in L_p(-\pi, \pi)$ ,  $1 \leq p \leq \infty$ , or  $C[-\pi, \pi]$  and its Poisson transform is*

$$f(u; r) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{(1-r^2)f(t+u)dt}{1-2r\cos t+r^2}, \quad 0 < r < 1, \quad (4)$$

*then  $\liminf_{r \rightarrow 1} (1-r)^{-1} \|f(u) - f(u; r)\| = 0$  implies that  $f(u)$  is a constant. Further  $\|f(u) - f(u; r)\| \leq \log(1/r) \|\tilde{f}'(u)\|$  whenever  $\tilde{f}'(u)$  belongs to the space.*

If  $E$  is a separable space or, more generally, if the set  $\{T_s x\}$ ,  $0 < s < \infty$ , is separable for every fixed  $x \in E$ , and if  $\omega(h)$  is a given modulus of continuity, then it is possible to find elements  $y \in E$  such that  $\|(T_h - I)y\| \leq C\omega(h)$ . For this purpose it is enough to choose a numerically valued function  $K(s)$  such that (i)  $K(s) \in L(0, \infty)$ , (ii)  $K(0) = 0$  and (iii)  $|K(s+h) - K(s)| \leq \omega(h)$  for  $0 \leq s < \infty$ , and to take

$$y = \int_0^\infty K(s) T_s x ds \quad (5)$$

where the integral exists in the sense of Bochner.<sup>3</sup>

3. In the preceding note<sup>1</sup> a representation theorem was announced [formula (2.5) compared with footnote 7]. The assumptions of this theorem are: (i)  $T_s$  is weakly measurable, (ii)  $\|T_s\| \leq 1$ , (iii)  $E$  is separable and (iv)  $\Sigma_s T_s(E)$  is dense in  $E$ . These assumptions imply, as was stated in our note, that (1)  $\|(T_h - I)x\| \rightarrow 0$  with  $h$  for all  $x$ , and (2)  $D(A)$  is dense in  $E$ . Now, conversely, if (1) is satisfied, then  $T_s$  is strongly continuous on the right for  $s \geq 0$  if  $T_0 = I$ . This implies (i) and (iv) and also that the set  $\{T_s x\}$ ,  $0 \leq s$ , is separable for every fixed  $s$ . It therefore turns out that (1) + (ii) are sufficient hypotheses for the representation theorem which can be formulated as follows:<sup>4</sup>

**THEOREM 3.** *If  $\|T_s\| \leq 1$  for  $s > 0$  and  $\|(T_h - I)x\| \rightarrow 0$  with  $h$  for all  $x$ , then  $\exp(sA_h)$  is a bounded linear transformation such that (1)  $\|\exp(sA_h)\| \leq 1$ ,  $s \geq 0$ ,  $h > 0$ , and (2)*

$$\lim_{h \rightarrow 0} \|\exp(sA_h)x - T_s x\| = 0, \quad (6)$$

uniformly in  $s$ ,  $0 \leq s \leq \omega < \infty$ .

The proof of (1) follows from

$$\exp(sA_h) = \exp(-s/h) \exp((s/h)T_h).$$

The proof of (2) may be based upon the well-known fact that

$$\lim_{t \rightarrow \infty} e^{-t} \sum \frac{t^n}{n!} = 1 \quad (7)$$

if the summation is extended merely over the terms for which  $(1 - \delta)t \leq n \leq (1 + \delta)t$ ,  $\delta$  fixed positive, while the summation over the complementary values of  $n$  gives the limit zero. We have

$$\exp(sA_h)x - T_s x = \exp(-s/h) \sum_0^\infty \frac{1}{n!} \left(\frac{s}{h}\right)^n [T_{nh}x - T_s x]. \quad (8)$$

If  $x \in D(A)$ ,  $T_s x$  is absolutely continuous in  $s$ . We can then choose a  $\delta$  such that  $\|T_{nh}x - T_s x\| \leq \epsilon \|x\|$  for  $(1 - \delta)s \leq nh \leq (1 + \delta)s$ , where  $\delta$  is independent of  $s$  in  $0 \leq s \leq \omega$ . Outside of this  $n$ -range we have  $\|T_{nh}x - T_s x\| \leq 2 \|x\|$ . Hence the right-hand side tends strongly to zero with  $h$ . This proves (2) for  $x \in D(A)$ . But  $D(A)$  is dense in  $E$  and  $\exp(sA_h)$  is uniformly bounded with respect to  $h$ . Hence (2) holds for all  $x \in E$ .

A special case is the following generalization of Taylor's theorem.<sup>6</sup>

**THEOREM 4.** *If  $f(u)$  is uniformly continuous in  $0 \leq u < \infty$ , then*

$$f(u + s) = \lim_{h \rightarrow 0} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{s}{h}\right)^n \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} f(u + kh), \quad (9)$$

where the limit exists uniformly with respect to  $u$  for all  $u$  and uniformly with respect to  $s$  for  $0 \leq s \leq \omega$ . If instead  $f(u) \in L_p(0, \infty)$ ,  $p$  fixed,  $1 \leq p < \infty$ , then the limit exists in the sense of convergence in the mean of order  $p$ .

4. The relations between the spectral properties of  $A$  and of  $T_s$  are of importance. We use the symbols  $S(U)$ ,  $PS(U)$ ,  $CS(U)$  and  $RS(U)$  to denote the spectrum of  $U$  and its point, continuous and residual components. The resolvent of  $U$  is denoted by  $R(\lambda, U)$ .

**THEOREM 5.** *Under the assumptions of Theorem 3,  $\alpha \in S(A)$  implies  $e^{as} \in S(T_s)$  for all  $s > 0$ . In particular,  $e^{as} \in PS(T_s)$  if  $\alpha \in PS(A)$ .*

The proof follows, for instance, from the identity

$$(T_s - e^{as}I)x = \int_0^s e^{a(s-t)}(A - \alpha I)T_t x dt, \quad x \in D(A). \quad (10)$$

This formula shows that  $(A - \alpha I)x = 0$  implies  $(T_s - e^{as}I)x = 0$  which takes care of the point spectrum. If  $\alpha \in CS(A)$ ,  $\|(A - \alpha I)x\|$  is not bounded away from zero on the unit sphere and the same is obviously true for  $\|(T_s - e^{as}I)x\|$ . Thus  $R(e^{as}, T_s)$  either does not exist or is unbounded.

Finally, if  $(A - \alpha I)$  maps  $D(A)$  upon a space  $E_\alpha$  non-dense in  $E$ , then  $(T_s - e^{\alpha s} I)$  maps  $E$  upon a subspace of the closure of  $E_\alpha$ . Thus if  $R(e^{\alpha s}, T_s)$  exists, its domain of definition cannot be dense in  $E$  and  $e^{\alpha s}$  belongs either to  $PS(T_s)$  or  $RS(T_s)$ .

Theorem 5 admits of a limited converse. If it is known that  $e^{\alpha s} \in S(T_s)$  for  $s = s_1$  and  $s_2$  where  $s_1/s_2$  is irrational, and if, in addition,  $e^{\alpha s_1}$  and  $e^{\alpha s_2}$  are spectral values of the same type in a certain narrow sense (for instance, if both belong to the point-spectrum and have a common characteristic element), then  $e^{\alpha s} \in S(T_s)$  for all  $s > 0$  and  $\alpha \in S(A)$ .

The spectral behavior of  $A$  and the approximation of  $I$  by  $T_h$  are closely related. The following theorem indicates such a relation.

**THEOREM 6.** *If  $\|T_h - I\| \rightarrow 0$  with  $h$  and  $\|T_h - I\| < 1 - 1/e$  for  $h < \rho$ , and if  $|\alpha| > 1/\rho$ , then  $\|(A - \alpha I)x\|$  is bounded away from zero on the intersection of  $D(A)$  with the unit-sphere. In particular,  $PS(A)$  and  $CS(A)$  are located inside the circle  $|\alpha| = 1/\rho$ .<sup>6</sup>*

The assumption  $\|T_h - I\| \rightarrow 0$  is more than sufficient to ensure the validity of (10). Denoting  $\int_0^s |\exp[\alpha(s-t)]| dt$  by  $M(\alpha, s)$ , we then get

$$M(\alpha, s) \|(A - \alpha I)x\| \geq |1 - e^{\alpha s}| \|x\| - \|(T_s - I)x\|. \quad (11)$$

For  $s = 1/|\alpha| < \rho$ , the right-hand side is positive and the theorem follows.

<sup>1</sup> "Representation of One-Parameter Semi-Groups of Linear Transformations," these PROCEEDINGS, 28, 175-178 (1942).

<sup>2</sup> "Notes on Linear Transformations," *Trans. Amer. Math. Soc.*, 39, 131-153 (1936).

<sup>3</sup> Compare I. Gelfand's construction of elements in  $D(A)$  in "On One-Parametrical Groups of Operators in a Normed Space," *Compt. Rend. Acad. Sci. U. R. S. S.*, 25, 713-718 (1939).

<sup>4</sup> For Theorems 3-6 the author had originally found proofs based upon the resolvent theory sketched in the earlier note. This method has the advantage of great general applicability, but it is fairly complicated and in special cases stronger results are obtained more easily by direct methods such as those indicated in the present note. The proof of Theorem 3 given here is an adaptation to the abstract case of an argument suggested by Professor G. Szegő for the proof of Theorem 4. A still simpler and more direct proof has just been found by Professor Nelson Dunford.

<sup>5</sup> Professor Dunford has called my attention to the fact that Theorem 4 gives a new and fairly direct proof of Weierstrass' approximation theorem.

<sup>6</sup> This theorem suggests strongly that  $RS(A)$  is also bounded and, as a consequence, that  $A$  is a bounded operator. My methods do not yield this result, but it has been proved very elegantly by Dunford whose proofs will appear elsewhere.

# GENERALIZATION OF SUNDMAN'S FUNDAMENTAL EQUALITY TO THE CASE OF MORE THAN THREE BODIES

BY GODOFREDO GARCÍA

NATIONAL UNIVERSITY OF SAN MARCOS, LIMA

Communicated August 27, 1942

In agreement with my eminent friend, Professor George D. Birkhoff, in considering that the physical probability of simple or multiple collision is zero and that it is very important from the physical point of view to treat the general typical case of movements without collision, I am led to investigate the possible generalization of the celebrated equality of Sundman to the case of the problem of  $n$  bodies.

Let  $P_i [i = 0, 1, 2, \dots, (n - 1)]$  be  $n$  points of masses  $m_i [i = 0, 1, 2, \dots, (n - 1)]$  which move according to the law of Newton; the total mass of the system is  $M = \sum_{i=0}^{n-1} m_i$  and the potential function is expressed by

$$U_0 = k \sum_{i=0}^{n-1} \frac{m_i m_j}{r_{ij}}, \quad i \neq j. \quad (1)$$

The vectorial equation of movement for any one of the  $n$  bodies is

$$m_i \frac{d^2(P_i - G)}{dt^2} = m_i \frac{d^2 \vec{L}_i}{dt^2} = \vec{F}_i = \text{grad}_i U_0, \quad i = 0, 1, 2, \dots, (n - 1). \quad (2)$$

Representing by

$$D_{ij} = P_i - P_j = \vec{\Phi}_{ij} r_{ij}, \quad i \neq j = 0, 1, 2, \dots, (n - 1), \quad (3)$$

the vectors which correspond to the material points  $P_i$  considered by pairs, then

$$\text{grad}_i U_0 = \sum_{j \neq i} \frac{\partial U_0}{\partial r_{ij}} \frac{P_i - P_j}{r_{ij}}, \quad i = 0, 1, 2, \dots, (n - 1). \quad (4)$$

Supposing the center of gravity  $G$  to be at the origin, the initial conditions are given by the equations

$$\sum_0^{n-1} m_i (P_i - G) = \sum_0^{n-1} m_i \vec{P}_i' = 0. \quad (5)$$

Representing by  $\vec{f}$  the constant vector of areas, we have

$$\sum_0^{n-1} (P_i - G) \wedge m_i \vec{P}_i' = \vec{f}. \quad (6)$$

The kinetic energy of the system formed by the  $n$  bodies taken in pairs is expressed by

$$T = \frac{1}{4M} \sum_0^{n-1} m_i m_j (\vec{P}_i' - \vec{P}_j') \times (\vec{P}_i' - \vec{P}_j') \quad (7)$$

which is transformed at once into the following:

$$T = \frac{1}{2} \sum_0^{n-1} m_i \vec{P}_i' \times \vec{P}_i' \quad (8)$$

since we know that

$$T = U_0 - K_0. \quad (9)$$

The expression of Lagrange for the moment of inertia of the  $n$  bodies is

$$S = R^2 = \frac{1}{2M} \sum_0^{n-1} m_i m_j (P_i - P_j) \times (P_i - P_j) \quad (10)$$

or

$$S = R^2 = \frac{1}{2M} \sum_0^{n-1} m_i m_j \times \{ [(P_i - G) - (P_j - G)] \times [(P_i - G) - (P_j - G)] \} \quad (11)$$

which, taking account of (5), reduces to

$$S = R^2 = \sum_0^{n-1} m_i (P_i - G) \times (P_i - G). \quad (12)$$

The first derivative of (12) is

$$S' = \frac{1}{M} \sum_0^{n-1} m_i m_j (P_i - P_j) \times (\vec{P}_i' - \vec{P}_j') \quad (13)$$

which is transformed into

$$S' = 2 \sum_0^{n-1} (P_i - G) \times m_i \vec{P}_i'. \quad (14)$$

The second derivative of (12) is

$$S'' = \frac{1}{M} \sum_0^{n-1} m_i m_j \{ (P_i - P_j) \times (\vec{P}_i'' - \vec{P}_j'') + (\vec{P}_i' - \vec{P}_j') \times (\vec{P}_i' - \vec{P}_j') \} \quad (15)$$

which becomes the following:

$$S'' = 2 \left\{ \sum_0^{n-1} (P_i - G) \times m_i \vec{P}_i'' + \sum_0^{n-1} m_i \vec{P}_i' \times \vec{P}_i' \right\}. \quad (16)$$

Making

$$\sum_0^{n-1} (P_i - G) \times m_i \vec{P}_i' = W \quad (17)$$

and observing that the function  $U_0$  is homogeneous and of dimensions  $-1$  in the distance which is the modulus of the vector  $(P_i - G)$ , one deduces

$$\sum_0^{n-1} (P_i - G) \times m_i \vec{P}_i' = \sum_0^{n-1} (P_i - G) \times \text{grad}_i U_0 = -U_0 \quad (18)$$

and taking account of (17) one has

$$S' = 2W \quad (19)$$

$$S'' = 2[-U_0 + 2T] = 2[U_0 - 2K_0]. \quad (20)$$

The function of Sundman is

$$H = RR'^2 + 2K_0R + \frac{f^2}{R} \quad (21)$$

or

$$H = \left[ \frac{S'^2}{4} + 2K_0S + f^2 \right]. \quad (22)$$

The derivative will be

$$H' = \left[ S \left( \frac{S''}{2} + K_0 \right) - \frac{1}{8} S'^2 - \frac{f^2}{2} \right] \frac{S'}{S^{3/2}} \quad (23)$$

and as

$$S'' = 4T - 2U_0 \quad (24)$$

and furthermore one has

$$\frac{S''}{2} = 2T - U_0 = U_0 - 2K_0 \quad (25)$$

from which one deduces

$$T = U_0 - K_0 = \frac{S''}{2} + K_0, \quad (26)$$

consequently the function of Sundman may be written:

$$H = [W^2 + 2K_0S + f^2] S^{-1/2} \quad (27)$$

and its derivative will be

$$H' = \left[ S(U_0 - K_0) - \frac{W^2 + f^2}{2} \right] \frac{S'}{S^{3/2}} \quad (28)$$

which is the fundamental equality of Sundman, extended to  $n$  bodies,



*GENERALIZATION OF THE INEQUALITY OF SUNDMAN TO  
THE CASE OF MORE THAN THREE BODIES AND TO THE CASE  
OF A GRAVITATIONAL GAS*

BY GODOFREDO GARCÍA

NATIONAL UNIVERSITY OF SAN MARCOS, LIMA

Communicated August 27, 1942

The present note is the continuation of the preceding; we proceed to demonstrate that the celebrated inequality of Sundman extends to the general case of  $n$  bodies and even to a gravitational gas such as a cloud or gaseous nebula.

Let us seek then the maximum or minimum of the function  $W$  for given  $\vec{P}_i'$ , or, as well, of  $S'$ , for accomplishing which we will avail ourselves of the method of multipliers of Lagrange.

Let the said multipliers be the vectors  $\vec{\lambda}$  and  $\vec{\rho}$  and the scalar  $\theta$ ; then with the equations (5), (6), (8) and (17) of Note I, we will form the equation

$$\begin{aligned} \vec{\lambda} \times \sum_0^{n-1} m_i \vec{P}_i' + \vec{\rho} \times \left[ \sum_0^{n-1} (P_i - G) \wedge m_i \vec{P}_i' - \vec{f} \right] + \\ \frac{\theta}{2} \left[ \sum_0^{n-1} m_i \vec{P}_i' \times \vec{P}_i' - U_0 + K_0 \right] + \\ \sum_0^{n-1} (P_i - G) \times m_i \vec{P}_i' - W = 0. \quad (29) \end{aligned}$$

Differentiating the function  $W$  with respect to  $\vec{P}_i'$  there are obtained the following  $n$  equations of structure:

$$\vec{\lambda} + \vec{\rho} \wedge (P_i - G) + \theta \vec{P}_i' + (P_i - G) = 0. \quad (30)$$

Multiplying them by  $m$  and summing, one has  $\vec{\lambda} = 0$  and (30) is reduced to

$$(P_i - G) + \vec{\rho} \wedge (P_i - G) + \theta \vec{P}_i' = 0. \quad (31)$$

Multiplying vectorially the equation (31) by  $m_i(P_i - G)$  and summing gives

$$\sum_0^{n-1} m_i [\vec{\rho} \wedge (P_i - G)] \wedge (P_i - G) - \theta \sum_0^{n-1} (P_i - G) \wedge m_i \vec{P}_i' = 0 \quad (32)$$

which is transformed into

$$\sum_0^{n-1} m_i [\vec{\rho} \times (P_i - G)] (P_i - G) - \vec{\rho} \sum_0^{n-1} m_i (P_i - G) \times (P_i - G) - \theta \sum_0^{n-1} (P_i - G) \wedge m_i \vec{P}_i' = 0. \quad (33)$$

Taking account of (6) and (12) one has

$$\sum_0^{n-1} m_i [\vec{\rho} \times (P_i - G)] (P_i - G) - \vec{\rho} S - \theta f = 0. \quad (34)$$

The most general structure of the vector  $\vec{\rho}$  which verifies the previous equation (34) is

$$\vec{\rho} = A(P_i - G) \wedge \vec{P}_i' + B(P_i - G) \quad (35)$$

where  $A$  and  $B$  are two indeterminates, and where it will be necessary to effect the determination of  $A$ . For doing this one substitutes  $\vec{\rho}$  in (31) and thus obtains the equation

$$(P_i - G) + A[(P_i - G) \wedge \vec{P}_i' \wedge (P_i - G)] + \theta \vec{P}_i' = 0 \quad (36)$$

or

$$[1 - A(P_i - G) \times \vec{P}_i'](P_i - G) + [\theta + A(P_i - G) \times (P_i - G)] \vec{P}_i' = 0 \quad (37)$$

which is verified under the conditions:

$$1 - A(P_i - G) \times \vec{P}_i' = 0, \quad \theta + A(P_i - G) \times (P_i - G) = 0 \quad (38)$$

$$A = \frac{1}{(P_i - G) \times \vec{P}_i'} \quad (39)$$

$$\theta = -A(P_i - G) \times (P_i - G) \quad (40)$$

$$\theta = - \frac{(P_i - G) \times (P_i - G)}{(P_i - G) \times \vec{P}_i'} = - \frac{\sum_0^{n-1} m_i (P_i - G) \times (P_i - G)}{\sum_0^{n-1} (P_i - G) \times m_i \vec{P}_i'} = - \frac{S}{W}. \quad (41)$$

Since

$$\vec{M}_\rho = A \sum_0^{n-1} (P_i - G) \wedge m_i \vec{P}_i' = \vec{A}f \quad (42)$$

one deduces

$$A = \frac{M}{W} \quad (43)$$

$$M\theta = -AS \quad (44)$$

and then

$$\vec{\rho} = \frac{\vec{f}}{W} = -\frac{\theta}{S} \vec{f}. \quad (45)$$

This structure of the vector  $\vec{\rho}$  satisfies the equation (34).

Multiplying (31) by the vector  $m_i \vec{P}_i$  and summing one has

$$\sum_0^{n-1} (P_i - G) \times m_i \vec{P}_i' + \vec{\rho} \times \sum_0^{n-1} (P_i - G) \wedge m_i \vec{P}_i' + \theta \sum_0^{n-1} m_i \vec{P}_i' \times \vec{P}_i' = 0 \quad (46)$$

$$W + \vec{\rho} \times \vec{f} + 2\theta(U_0 - K_0) \quad (47)$$

or

$$W + \theta \left[ \frac{\vec{\rho} \times \vec{f}}{\theta} + 2(U_0 - K_0) \right] = 0. \quad (48)$$

By the equation (45) one has also:

$$\frac{\vec{\rho} \times \vec{f}}{\theta} = -\frac{f^2}{S} \quad (49)$$

$$W\theta = -S \quad (50)$$

and multiplying (48) by  $W$  one finds

$$W^2 + f^2 - 2S(U_0 - K_0) \leq 0 \quad (51)$$

which corresponds to the maximum value of  $W$ .

The equations (19) and (28) enable us to write the preceding equation (51) in the following form:

$$\frac{S'^2}{4} + f^2 - 2S \left[ \frac{S''}{2} + K_0 \right] \leq 0 \quad (52)$$

or

$$S'' - \frac{S'^2}{4S} + 2K_0 \geq \frac{f^2}{S} \quad (53)$$

and as  $R^2 = S$  there is obtained

$$2RR'' + R'^2 + 2K_0 \geq \frac{f^2}{R^2} \quad (54)$$

which demonstrates that the natural extension of the inequality of Sundman is fulfilled in the problem of  $n$  bodies, attracting according to the law of Newton. Substituting then in (28), from

$$W^2 \leq -f^2 + 2S(U_0 - K_0) \quad (55)$$

there results

$$H' \geq 0 \quad (56)$$

$$H \leq 2U_0S^{1/2}. \quad (57)$$

*The Gravitational Gas.*—In the case of a gravitational gas one can, for instance, consider the masses as equal:

$$m_i = m_j = m, \quad i \neq j.$$

Furthermore we will write

$$S_i = (P_i - G) \times (P_i - G) \quad (58)$$

$$\vec{f}_i = (P_i - G) \wedge \vec{P}_i' \quad (59)$$

so that

$$S = \sum_0^{n-1} n_m S_i = \sum_0^{n-1} n_m (P_i - G) \times (P_i - G) \quad (60)$$

$$\vec{f} = \sum_0^{n-1} n_m \vec{f}_i = \sum_0^{n-1} n_m (P_i - G) \wedge \vec{P}_i'. \quad (61)$$

Setting

$$n_m = \delta_i \Delta v_i \quad (62)$$

where  $\delta_i$  is the density and  $\Delta v_i$  the volume of the elements,

$$S = \sum_0^{n-1} \delta_i S_i \Delta v_i = \sum_0^{n-1} \delta_i (P_i - G) \times (P_i - G) \Delta v_i \quad (63)$$

$$\vec{f} = \sum_0^{n-1} \delta_i \vec{f}_i \Delta v_i = \sum_0^{n-1} \delta_i (P_i - G) \wedge \vec{P}_i' \Delta v_i \quad (64)$$

$$S' = \sum_0^{n-1} \delta_i S_i' \Delta v_i = 2 \sum_0^{n-1} \delta_i (P_i - G) \times \vec{P}_i' \Delta v_i \quad (65)$$

$$S'' = \sum_0^{n-1} \delta_i S_i'' \Delta v_i = 2 \left[ \sum_0^{n-1} \delta_i \vec{P}_i' \times \vec{P}_i' \Delta v_i + \sum_0^{n-1} \delta_i (P_i - G) \times \vec{P}_i'' \Delta v_i \right] \quad (66)$$

$$\sum_0^{n-1} \delta_i S_i'' \Delta v_i - \frac{\left[ \sum_0^{n-1} \delta_i S_i' \Delta v_i \right]^2}{4 \sum_0^{n-1} \delta_i S_i \Delta v_i} + 2K_0 \geq \frac{\left[ \sum_0^{n-1} \delta_i \vec{f}_i \Delta v_i \right]^2}{\sum_0^{n-1} \delta_i S_i \Delta v_i} \quad (67)$$

Passing to the limit one has

$$\int_v \delta S_p'' dv - \frac{[\int_v \delta S_p' dv]^2}{4 \int_v \delta S_p dv} + 2K_0 \geq \frac{[\int_v \delta \vec{f}_p dv]^2}{\int_v \delta S_p dv} \quad (68)$$

which demonstrates that in the case of a gravitational gas there is fulfilled the inequality of Sundman.

While Sundman demonstrated that for a given initial state, position and velocities of the three bodies can be determined an inferior limit for the diameter for all time, of which result Birkhoff showed the true significance, I intend to consider later analogous questions for the general case of  $n$  bodies and even for a gravitational gas.<sup>1</sup>

<sup>1</sup> My colleague, Professor Alfred Rosenblatt, studied such gravitational gases as long ago as 1926, but did not publish his very interesting main result analogous to Sundman's Theorem I. This will appear shortly in his paper in the *Am. Jour. Math.*, "On the Movement of a Cosmic Cloud of Finite Mass and Dimensions Which Is Only Subjected to the Newtonian Law of Gravitation."

*ON THE ANAPHASE MOVEMENT OF CHROMOSOMES*

BY M. M. RHOADES AND HILDA VILKOMERSON

DEPARTMENT OF BOTANY, COLUMBIA UNIVERSITY

Communicated August 22, 1942

The kinetic properties of chromosomes are controlled by the centromere or kinetochore. In somatic mitoses it is the centromeres which become oriented on the equatorial plate while the two arms of each chromosome may extrude from the spindle figure into the cytoplasm. At metaphase a thread-like structure, descriptively known as the spindle fiber, can be seen in properly fixed and stained material extending from the centromere to the pole. The centric region plays a decisive rôle in congression, orientation on the spindle and anaphase separation. Acentric fragments behave in an irregular fashion. The centric region becomes "attached" to the spindle and leads the way to the pole at anaphase with the arms of the chromosome apparently being moved as passive bodies. A sub-terminal centromere imparts a J-shaped appearance to an anaphase chromosome while a chromosome with a median centromere assumes a V-shape. In either case the centromere (i.e., centric region) is at the apex of the configuration.

At the first meiotic metaphase the bivalent chromosomes undergo congression and orientation. When the bivalent (tetrad) moves upon the spindle the two homologous centromeres become cooriented and lie symmetrically on either side of the equatorial plane and directed toward opposite poles. The two homologous centric regions lead the way to opposite poles in anaphase disjunction. Inasmuch as two chromatids are attached to each centromere, the disjoining dyads have the appearance of double V's or J's, depending upon the position of the centromere in the chromosome. The double V's or J's characterizing the first anaphase are later transformed into X-shaped configurations by the repulsion between the four constituent arms. The undivided centromere lies at the center of the X and holds the two chromatids together until the second anaphase. The dyads at MII have their undivided centromeres lying on the equatorial plate as in a somatic mitosis. The effective split of the centromere occurs and the two chromatids comprising each dyad pass to opposite poles, with the centric region advancing foremost. In maize there is no relational coiling at the second division to delay disjunction. This fact coupled with the marked contraction of the chromosome arms at MII leads to the two chromatids moving somewhat parallelly to the poles, and the pronounced V's and J's characteristic of other mitoses are not always seen, although it is clear that the centric region is in the front. The above outline is true

for maize chromosomes, and with minor exceptions holds for those plants and animals possessing chromosomes with localized centromeres. Recently Hughes-Schrader and Ris<sup>1</sup> have shown that the hemipterous insects have a diffuse type of spindle attachment region and consequently in these insects the movement of the chromosomes follows another pattern. There is then for chromosomes no universality in type of spindle behavior. This is emphasized by the following preliminary report of an anomalous situation in maize.

The unorthodox behavior is limited to the two meiotic divisions. The somatic mitoses are normal. Studies of meiosis have been limited to microsporogenesis. The first meiotic division is normal up to metaphase, when the bivalents congress upon the spindle figure. The pachytene chromosomes are of normal appearance with clearly defined centromeres. The bivalents become oriented on the spindle in a regular manner. The first indication of an unusual behavior occurs when structures similar to the primary centric region arise from distal portions of the chromosomes before the beginning of anaphase. These newly arisen structures will be tentatively called secondary centric regions inasmuch as they, like the primary centric region, become attached to the spindle and affect anaphase movements of the chromosomes. These secondary centric regions move poleward more rapidly than does the primary centric region, so that the distal ends of the chromosome, instead of facing the center of the spindle as is commonly true at anaphase, are pulled ahead and, overtaking the true centric region, come to lead the way to the poles. These secondary centric regions may be formed by one or more of the four arms comprising each dyad. The activity of these regions superimposed on the anaphase movement due to the primary centric region produces great complexity in the types of configurations observed. At the end of AI ten dyads usually are found at each pole, there being a surprisingly high regularity in disjunction.

Cytological conditions in the second division are much more favorable for observation and diagrammatic configurations are often found. Second prophase stages do not noticeably differ from normal but the onset of metaphase occurs before the usual contraction of the chromosomes has taken place. However, the somewhat extended dyads usually become oriented on the metaphase plate with the undivided centromere lying on the equatorial plate. Before the centromere splits and a normal anaphase separation is initiated, secondary centric regions again arise from or near the distal ends of the chromosomes. These new centric regions become attached to the spindle and move rapidly poleward, with the result that the chromosome arms become greatly attenuated if their proximal portions were anchored on the metaphase plate by the still undivided centromere (see Fig. 1). The centromere eventually divides and the monads pass to the poles.

When both ends of the same chromosome form secondary centric regions the chromosome literally backs into the pole with the apex (representing the centromere) of the V- or J-shaped chromosome pointing toward the equatorial plate while the two distal ends lead the way. Often only the distal end of one of the two arms forms a centric region, in which case the chromosome becomes an extended rod-shaped element. Occasionally both arms of the same chromosome may form secondary centric regions which are directed toward opposite poles, thus forming a chromosome bridge. Usually one of the two opposing forces prevails and the chromosome passes to one of the poles but infrequently the chromosome becomes suspended on the spindle with a consequent breaking of the chromosome at late anaphase. It is somewhat surprising that in both the first and second meiotic divisions regular disjunction usually occurs and that plants exhibiting this decidedly aberrant behavior are very fertile.

The formation of these secondary centric regions is limited to those plants having an abnormal type of chromosome 10<sup>2, 3</sup> with extra chromatin near the distal end of the long arm. In plants homozygous for this abnormal tenth chromosome the frequency is high for the formation of these secondary centric regions, while it is much less in plants heterozygous for this chromosome. Sister plants homozygous for a normal chromosome 10 had a completely orthodox behavior.

It is obvious that the unparalleled behavior reported here is of great interest to current theories on cell mechanics, especially that of the kinetic movement of chromosomes. A detailed account will be published elsewhere.

**Summary.**—In maize the primary centric region representing the localized centromere is responsible for the kinetic movement of the chromosome in anaphase. The concentration of the kinetic forces produces J- or V-shaped configurations in anaphase. Plants carrying an abnormal type



FIGURE 1

Second metaphase showing precocious poleward movement of secondary centric regions. The functionally undivided centromeres of some of the dyads are oriented on the metaphase plate. In some cells the secondary centric regions reach the periphery of the cell before division of the centromere occurs resulting in an extreme attenuation of the chromosome arms.



of chromosome 10 exhibited a unique behavior in that centric regions were formed by portions of the chromosome other than the centromere.

<sup>1</sup> Hughes-Schrader, S., and Ris, H., *Jour. Exp. Zool.*, **87**, 429-456 (1941).

<sup>2</sup> Rhoades, M. M., *Genetics*, **27**, 395-407 (1942).

<sup>3</sup> Longley, A. E., *Jour. Agric. Res.*, **56**, 177-195 (1938).

---

## HEAT-INDUCED TRIPLOIDY IN THE NEWT, *TRITURUS VIRIDESCENS*

BY GERHARD FANKHAUSER AND RITA CROTTA WATSON

DEPARTMENT OF BIOLOGY, PRINCETON UNIVERSITY

Communicated September 2, 1942

Heat and cold have both been widely used to induce polyploidy in plants.<sup>1, 2, 3, 4</sup> Depending on whether the treatment is applied during meiosis or during the early divisions of the zygote, triploid or tetraploid plants are produced. In general, the abnormal temperature seems to disturb the formation of the spindle and the normal separation of the daughter chromosomes at anaphase, probably through changes in the viscosity of the cytoplasm. Exposure during early prophase of meiosis may also affect the behavior of the chromosomes directly and prevent pachytene pairing,<sup>3, 4</sup> or completely suppress both meiotic divisions so that a diplotene nucleus changes directly into a pollen-grain nucleus.<sup>5</sup>

Similar effects of temperature have been described in some invertebrate animals. Refrigeration of normally parthenogenetic eggs of the brine shrimp, *Artemia salina*, may double the chromosome number from diploid to tetraploid through inhibition of the single, equational maturation division.<sup>6, 7</sup> Heat treatment of unfertilized eggs of the silkworm, *Bombyx mori*, induces parthenogenesis and also causes retention of the diploid chromosome number, sometimes with subsequent fusion of diploid cleavage nuclei to form tetraploid or partially tetraploid animals.<sup>8, 9, 10</sup>

Among vertebrates, spontaneous and experimentally induced polyploidy have been studied extensively in several species of salamanders, because of the ease with which the chromosome number of living young larvae may be determined in whole-mounts of the amputated tailtip.<sup>11</sup> Spontaneous deviations from the normal, diploid chromosome number occur rather frequently.<sup>12</sup> Among 1878 larvae of the newt, *Triturus viridescens*, which were examined from November, 1937, to August, 1942, 38 were found to possess various deviating chromosome numbers; the majority of these, 25 (1.33% of the total), were triploid.

Experimental triploidy was first induced in salamanders by refrigeration

of freshly fertilized eggs of *Triturus viridescens*.<sup>13, 14, 15</sup> Before insemination, which occurs during egg-laying, the amphibian egg has reached the metaphase of the second maturation division; it remains in this stage until fertilization has taken place. In the newt, late anaphase is reached about 30 minutes after fertilization.<sup>16</sup> The low temperature presumably suppresses the second maturation division and produces a diploid egg nucleus which fuses with a normal, haploid sperm nucleus to form a triploid cleavage nucleus. A cytological investigation of the events taking place in refrigerated eggs is in progress.

The range of low temperatures which will induce triploidy is considerable. Temperatures from 0° to +4°C. seem to be most effective; however,

TABLE 1  
SUMMARY OF RESULTS OF REFRIGERATION EXPERIMENTS WITH  
EGGS OF THREE SPECIES OF SALAMANDERS

Beginning of treatment: immediately or within a few minutes after fertilization.

Duration of treatment: 5 to over 24 hours.

Species	Temperature	Number of eggs treated	Number of larvae obtained	Chromosome Number			% 3N
				3N	2N	N	
<i>Triturus</i> <i>viridescens</i> <sup>1</sup> (2N = 22)	0° to +4.35°C.	466	228 (48.9%)	167 <sup>2</sup>	50	10	73.2
<i>Triturus</i> <i>pyrrhogaster</i> (2N = 24)	+1.5° to +2.5°C.	117	29 (24.8%)	13	11 <sup>2</sup>	4	44.8
<i>Amblystoma</i> <i>mexicanum</i> , <i>Axolotl</i> (2N = 28)	+1° to +3°C.	154	31 (20.1%)	25	5	1	80.6

<sup>1</sup> These figures include all experiments performed so far, except those involving temperatures above 4.35° that are not as effective in inducing triploidy.

<sup>2</sup> Plus one mixed hyperdiploid-triploid larva.

<sup>3</sup> Plus one hyperdiploid larva.

triploid larvae have also been obtained in considerable numbers following exposure of eggs to +5.5° and +6.38°C.<sup>16</sup> The upper and lower limits of the effective range have not yet been determined accurately. The duration of the treatment varied from 5 to over 24 hours; a few experiments of shorter duration produced diploid larvae only.<sup>14</sup> The percentage of triploid larvae varied in different series of experiments, largely because the eggs of individual females seem to react differently to the same treatment.

The refrigeration method has also been successfully applied to eggs of *Triturus pyrrhogaster*<sup>17</sup> and, more recently, in collaboration with Dr. Humphrey of the University of Buffalo, to axolotl eggs<sup>18</sup> (table 1). Although the total number of eggs treated in these two species was rather

small, the results indicate that specific differences exist, both in the rate of mortality during early stages of development and in the percentage of triploid individuals among the surviving larvae. It is also of interest that this treatment occasionally produces haploid larvae in all three species. The mechanism which leads to the formation of a haploid cleavage mitosis under cold treatment is not known at present.

In view of the effectiveness of heat treatment as a polyploidy-inducing agent in plants, experiments to test the influence of high temperatures on salamander eggs had been considered for some time. Following preliminary tests of the heat resistance of eggs of *Triturus viridescens* by several undergraduate students, the first extensive series of heat treatments were carried out by Rita Crotta Watson in the spring of 1942. The eggs were transferred immediately after laying to dishes kept in an incubator running at 34.2° to 37.2°C. Following the treatment which lasted for from 5 to over

TABLE 2  
RESULTS OF HEAT TREATMENT OF FRESHLY FERTILIZED EGGS OF  
*Triturus viridescens*

Beginning of treatment: immediately after laying.

Duration of treatment: 5 to over 50 minutes (see table 3).

	Temperature				Total
	34.2°C.	35.0°-35.5°C.	36.0°-36.8°C.	37.0°-37.2°C.	
No. of eggs treated	9	23	52	29	113
No. of larvae obtained	4 (44.4%)	10 (43.5%)	32 (62.5%)	16 (55.2%)	62 (54.9%)
Triploid	0	9 (90%)	25 (78.1%)	15 (93.8%)	49 (79%)

50 minutes (tables 2 and 3) the eggs were returned to water at room temperature and allowed to develop. The chromosome number of each surviving larva was determined from counts in the amputated tailtip.

The rate of mortality during early stages of development did not differ significantly at the various temperatures used (table 2). On the other hand, at any one of these temperatures, treatments lasting for over 19 minutes seemed to be increasingly harmful (table 3). The mortality among heat-treated eggs was highest during the blastula stage; in the refrigeration experiments, on the other hand, it reached its peak during early cleavage.

The four larvae which developed from eggs exposed to 34.2° were all diploid. Of the larvae obtained from treatments at 35 to 37.2°C. 84.4% were triploid; within this range, different temperatures seemed to be about equally effective in inducing triploidy (table 2). The duration of the treatment also had no clear-cut influence on the resulting percentage of

triploid larvae (table 3). The shortest treatments that produced triploid larvae were 5 minutes at 37.0°C., and 8 minutes at 36.0°C.

More experiments will be needed to determine the range of effective temperatures and lengths of treatments. Already it is evident that short heat treatments are at least as effective in inducing triploidy as long cold treatments, if not more so (compare tables 1 and 2). Furthermore, the rate of mortality among the heat-treated eggs may be slightly lower.

It is interesting to compare these results with those obtained by heat treatments of various species of plants during the early divisions of the zygote (cf. ref. 15, table 1). The temperatures used ranged from 38° to 45°C., and the duration of the treatment varied from 30 minutes to 48 hours. The yield of polyploids (predominantly tetraploids) varied from 0.25 to 8%. The greater effectiveness of heat treatments of amphibian eggs is probably determined largely by three factors: (1) the treatment

TABLE 3  
EFFECTS OF VARIATIONS IN DURATION OF HEAT TREATMENT AT  
35° TO 37.2°C.

	Duration, Minutes					
	5-9	10-19	20-29	30-39	40-49	Over 50
Number of eggs treated	7	47	24	13	10	3
Number of larvae obtained	5 (71.4%)	34 (72.3%)	12 (50%)	5 (38.5%)	2 (20%)	0
Triploid	5 (100%)	29 (85.3%)	8 (66.7%)	5 (100%)	2 (100%)	..

can be applied directly to the isolated egg cell; (2) it reaches the egg always in exactly the same stage of the mitotic cycle, the metaphase of the second maturation division; (3) the movements of the chromosomes are normally suspended until about 10 to 15 minutes after fertilization, with the chromatids still closely associated in pairs (dyads); the treatment thus merely prevents these movements from resuming their course instead of stopping an actively progressing mitosis.

Attempts to induce tetraploidy by refrigeration of newt eggs during the first cleavage mitosis have been a complete failure so far, probably because of the difficulties encountered in timing the treatment. The living egg is opaque and does not indicate in any way when it reaches the metaphase of the first division; moreover, the study of sections through preserved eggs has shown that the interval between fertilization and metaphase varies widely between individual eggs kept at the same temperature. It is possible that heat treatments, which are more easily controlled because of the short exposure that is needed, will be more successful.

*Summary.*—One hundred and thirteen freshly fertilized eggs of the newt, *Triturus viridescens*, were treated at temperatures ranging from 34.2° to 37.2°C., for from 5 to over 50 minutes, and raised at room temperature. Sixty-two (54.9%) of the treated eggs developed into larvae; 49 of these (79%) were found to be triploid, the rest diploid. The shortest heat treatments which produced triploid larvae were 5 minutes at 37.0°C., and 8 minutes at 36.0°C.

A comparison of the results with those obtained by refrigeration of eggs shows that short heat treatments are at least as effective in inducing triploidy as cold treatment lasting for 5 or more hours. Both heat and cold presumably suppress the second maturation division of the egg which is not completed normally until about one hour after fertilization.

- <sup>1</sup> Belling, J., *Jour. Hered.*, **16**, 463-464 (1925).
- <sup>2</sup> Randolph, L. F., *Proc. Nat. Acad. Sci.*, **18**, 222-229 (1932).
- <sup>3</sup> Sax, K., *Jour. Arnold Arboretum*, **17**, 153-159 (1936).
- <sup>4</sup> Sax, K., *Am. Jour. Bot.*, **24**, 218-225 (1937).
- <sup>5</sup> Barber, H. N., *Proc. Roy. Soc. London*, **128B**, 170-185 (1940).
- <sup>6</sup> Gross, F., *Naturwiss.*, **20**, 962-967 (1932).
- <sup>7</sup> Gross, F., *Z. Zellforschg. mikr. Anat.*, **23**, 522-565 (1935).
- <sup>8</sup> Hasimoto, H., *Bull. Imp. Sericult. Exp. Sta.*, **8**, 359-381 (1933).
- <sup>9</sup> Astaurov, B. L., "Artificial Parthenogenesis in the Silk Worm (*Bombyx mori* L.)," Moscow (1940).
- <sup>10</sup> Frolova, S. L., *Compt. Rend. (Doklady) Acad. Sci. URSS.*, **27**, 601-603 (1940).
- <sup>11</sup> Fankhauser, G., *Proc. Am. Philos. Soc.*, **79**, 715-739 (1938).
- <sup>12</sup> Fankhauser, G., *Proc. Nat. Acad. Sci.*, **27**, 507-512 (1941).
- <sup>13</sup> Fankhauser, G., and Griffiths, R. B., *Ibid.*, **25**, 233-238 (1939).
- <sup>14</sup> Griffiths, R. B., *Genetics*, **26**, 69-88 (1941).
- <sup>15</sup> Fankhauser, G., *Biol. Symposia*, **6**, 21-35 (1942).
- <sup>16</sup> Fankhauser, G., and Caroline Moore, *Jour. Morph.*, **68**, 347-385 (1941).
- <sup>17</sup> Fankhauser, G., R. Crotta and M. Perrot, *Jour. Exp. Zool.*, **89**, 167-181 (1942).
- <sup>18</sup> Fankhauser, G., and Humphrey, R. R., *Biol. Bull.* (in press).

## THE INCREASE OF B VITAMINS IN GERMINATING SEEDS\*

BY PAUL R. BURKHOLDER AND ILDA MCVEIGH

OSBORN BOTANICAL LABORATORY, YALE UNIVERSITY

Communicated September 8, 1942

The rapid synthesis of provitamins A and vitamin C in germinating seeds has been clearly demonstrated by many investigators in recent years.<sup>1, 2</sup> Much less is known, however, concerning the possible changes in the amounts of B vitamins during germination. Mung bean sprouts are known to contain vitamins of the B complex.<sup>3</sup> It has been reported that the B<sub>1</sub> of pea embryos increases during germination while the content

of the cotyledons decreases.<sup>4</sup> Moreover, it appears that the thiamine in young pea plants kept in the dark does not increase, whereas in the light the vitamin B<sub>1</sub> content of the leaves rises rapidly.<sup>5</sup> Sprouted garden peas have been found to produce also an appreciable quantity of riboflavin during their early development.<sup>6</sup> In view of the relative scarcity of information on the subject, it seemed desirable to make some assays for B vitamins in germinating seeds of a number of edible species.

*Materials and Methods.*—Ten kinds of plants, including Alpha barley, Victory oat, Cornell 34-53 corn, Tenmarq wheat, Bansei soy bean, Mung bean, Alaska garden pea, green eye pea, and large and small lima bean, were employed in this investigation of B vitamins in dry seeds and seedlings. The seeds were germinated at ordinary temperatures in sand cultures in the greenhouse or in liquid cultures in the laboratory. Tap water was supplied as required for normal growth. A few plants were grown on moist filter paper in large Petri dishes, one set being exposed to daylight and another kept in darkness.

After different periods of time whole seedlings were harvested and dried in an oven at 70°C. for about 12 hours. The material was ground fine in a glass mortar and preserved in the dry state over calcium chloride in a desiccator until the assays could be completed a few days later. Extracts of the dried materials were obtained by autoclaving each 0.5-g. sample in 20 ml. of 1 *N* H<sub>2</sub>SO<sub>4</sub> at 15 lbs. pressure for 45 minutes. Using a glass electrode, the reaction of the cooled solutions was adjusted to pH 5.0 by adding NaOH. The extracts containing finely suspended matter were made up with distilled water to a volume of 30 ml. and then filtered through Hirsch filters, using a small amount of washed asbestos. In all the work, glass-distilled water was employed, glassware was cleaned in chromic acid and washed thoroughly, and manipulations were carried out under dim Mazda light or in filtered red light until the riboflavin determinations were completed.

The methods used in assaying for riboflavin, biotin and niacin involved the use of *Lactobacillus casei*  $\epsilon$ , *Saccharomyces cerevisiae* F. B. and *Lactobacillus arabinosus* as indicators in microbiological tests described by Snell and Strong,<sup>7</sup> Snell, Eakin and Williams,<sup>8</sup> and Snell and Wright.<sup>9</sup> Thiamine activity was tested with the *Phycomyces* assay method.<sup>10</sup> Pyridoxine was assayed tentatively with a yeast which has been in use for some time in this laboratory, following the general procedure of Williams, Eakin and McMahan.<sup>11</sup> The amounts of the extracts to be used in making the tests were determined by preliminary trials, and appropriate aliquots were employed usually at several concentration levels for each vitamin assay so that growth of the indicator organism would fall within a suitable range of response. The resulting data are expressed as micrograms of vitamin per gram of dry matter.

In view of difficulties with the microbiological assay for riboflavin in cereals, as reported by several investigators,<sup>12, 13</sup> it seemed desirable to determine to what extent other factors might be influencing the growth of the test organism, *Lactobacillus casei*. It should be pointed out that starch is effectively hydrolyzed in the samples by autoclaving them in 1 *N* H<sub>2</sub>SO<sub>4</sub>. Experiments were conducted with photolyzed extracts of dormant and germinated barley in order to ascertain the extent to which interfering substances might modify the results of the assays. The plant materials were extracted in the usual way by autoclaving in 1 *N* H<sub>2</sub>SO<sub>4</sub> and filtering off the suspended matter. Then the liquid samples were adjusted with NaOH to pH 11 and exposed to strong Mazda light for 15 hours to destroy the riboflavin. The reaction was readjusted to pH 5.0 just before using the extracts in the tests. Series of tubes were then set up to contain the following amounts of riboflavin: none, 0.05 gamma, and 1.0 gamma per 10 ml. of culture medium. One milliliter of extract containing an equivalent of 17 mg. of dry barley, approximately the amount used in the assays for B<sub>2</sub>, was added to each tube. Uninoculated controls were run in the usual way. Growth of the bacteria was estimated turbidimetrically after 18 hours.

The results indicate that riboflavin was absent in the photolyzed extracts. At the 0.05 gamma level of B<sub>2</sub>, where this vitamin is presumed to be the growth factor in relative minimum, no increase in growth was observed in tubes containing additions of photolyzed extract as compared with control tubes containing the same amount of riboflavin but no extract; at the level of 1.0 gamma of B<sub>2</sub> per tube, where the vitamin is not a limiting factor, increased growth was noted especially in the tubes which had received extract of germinated barley. Inasmuch as our determinations of riboflavin were performed at the lower end of the dose-response curve, it is believed that the influence of interfering substances was kept at a minimum.

The results of the assays for pyridoxine are still in a preliminary state and a separate report is planned for this part of the work. Only a few tentative results with B<sub>6</sub> are therefore presented in this paper.

*Results.*—In some of our preliminary experiments (completed April 13 and April 30, 1942) seeds were germinated after sterilization for 30 minutes in calcium hypochlorite (10 g. in 140 ml. H<sub>2</sub>O). The seedlings were grown under aseptic conditions in moist Petri dishes, some being exposed to daylight while others were kept in darkness. A few of the results obtained with seedlings kept in darkness are shown in table 1. In nearly every instance there was a loss of vitamins in the soaked seed, probably resulting from leaching and the action of the chlorine. After 4 days of growth the niacin, riboflavin and biotin in nearly every case exceeded the initial concentrations present in the dry seeds. Assays for riboflavin in both light



and dark grown seedlings indicated increases of about the same order in the developing plants of both series.

The results of another series of assays (completed May 27, 1942) on seedlings sprouted for 5 days in sand culture are presented in table 2. In this experiment the non-sterilized seeds were planted directly in moist sand. Riboflavin, niacin and biotin show relatively large gains in the

TABLE 1

VITAMINS IN DRY SEEDS AND IN GERMINATED SEEDS STERILIZED IN  $\text{CaCl}_2(\text{OCl})$  SOLUTION.  
RESULTS EXPRESSED AS MICROGRAMS PER GRAM OF DRY MATTER

KIND OF SEED	TREATMENT	RIBOFLAVIN	NIACIN
Wheat	Dry	1.0	83
	Soaked	1.4	67
	4 Days old	2.4	72
Corn	Dry	1.5	20
	Soaked	0.5	16
	4 Days old	2.0	30
Soy bean	Dry	1.8	26
	Soaked	1.4	20
	4 Days old	3.6	37
Mung bean	Dry	1.2	30
	Soaked	1.2	29
	4 Days old	2.6	51

TABLE 2

VITAMIN CONTENT (MICROGRAMS PER GRAM) OF DRY SEEDS AND OF SEEDS GERMINATED FOR 5 DAYS IN SAND CULTURES

KIND OF SEED	RIBOFLAVIN		NIACIN		THIAMINE		BIOTIN	
	DRY	GERMIN.	DRY	GERMIN.	DRY	GERMIN.	DRY	GERMIN.
Barley	1.3	8.3	72	129	..	7.9	0.4	1.2
Corn	1.2	3.0	17	40	6.2	5.5	0.3	0.7
Oats	0.6	12.4	11	48	10.0	11.5	1.2	1.8
Soy bean	2.0	9.1	27	49	10.7	9.6	1.1	3.5
Lima (large)	1.0	2.0	15	29	6.7	5.0	0.1	0.1
Lima (small)	0.9	4.0	11	41	4.5	6.2	0.1	0.4
Green eye pea	1.8	9.7	20	60	11.0	12.0	0.4	1.1
Mung bean	1.2	10.0	26	70	8.8	10.3	0.2	1.0
Pea	0.7	7.3	31	32	7.2	9.2	...	0.5

germinated seeds. Little or no change was observed for the thiamine content in this set of plants. The results for pyridoxine were not entirely satisfactory because of interfering substances in the samples, but it can be said that the preliminary data indicate large increases during early germination.

The results of further experiments with corn and wheat seedlings grown in water culture in the laboratory are shown in figure 1. Appreciable increases in the concentration of riboflavin, niacin and biotin are apparent



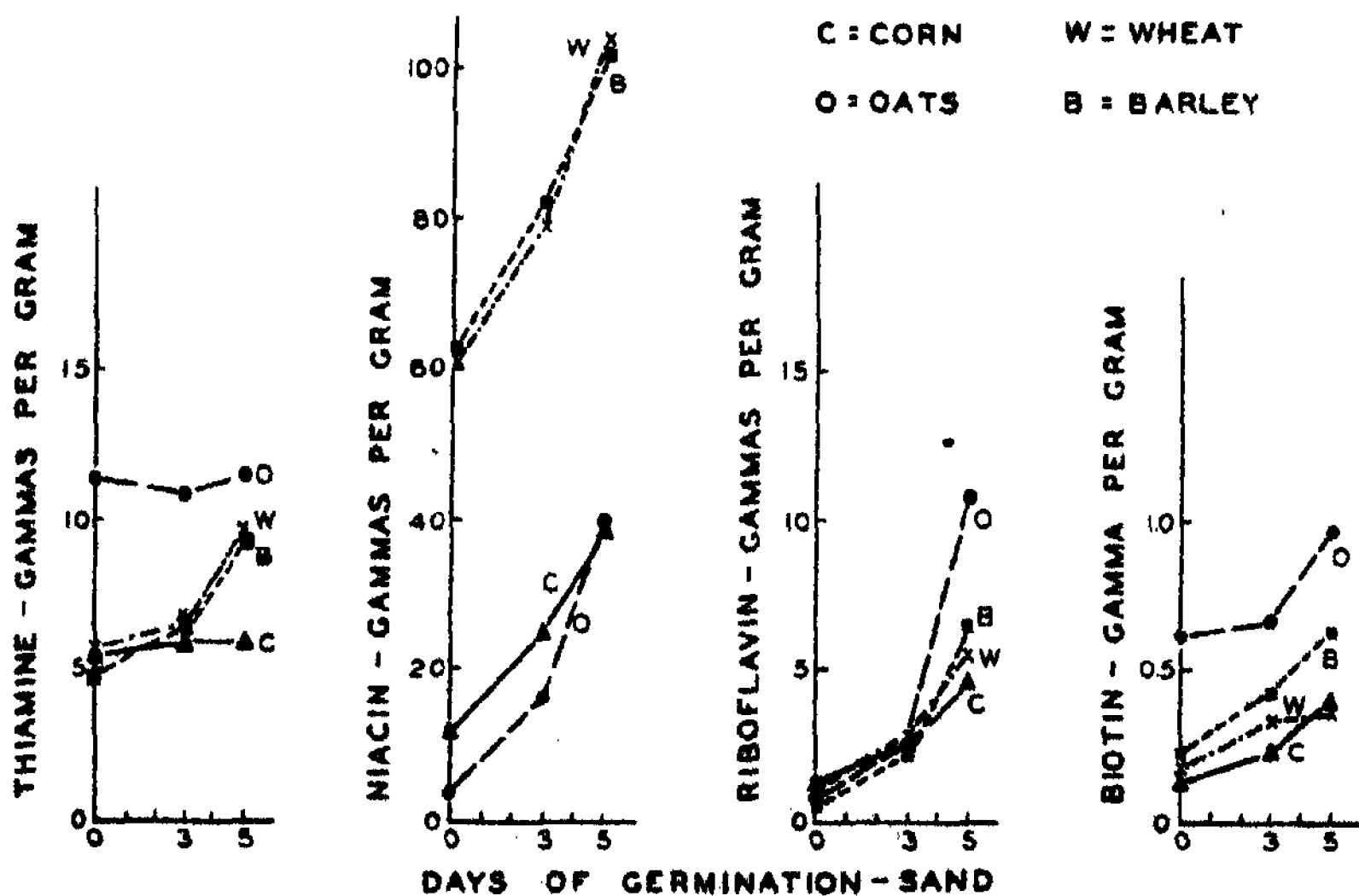
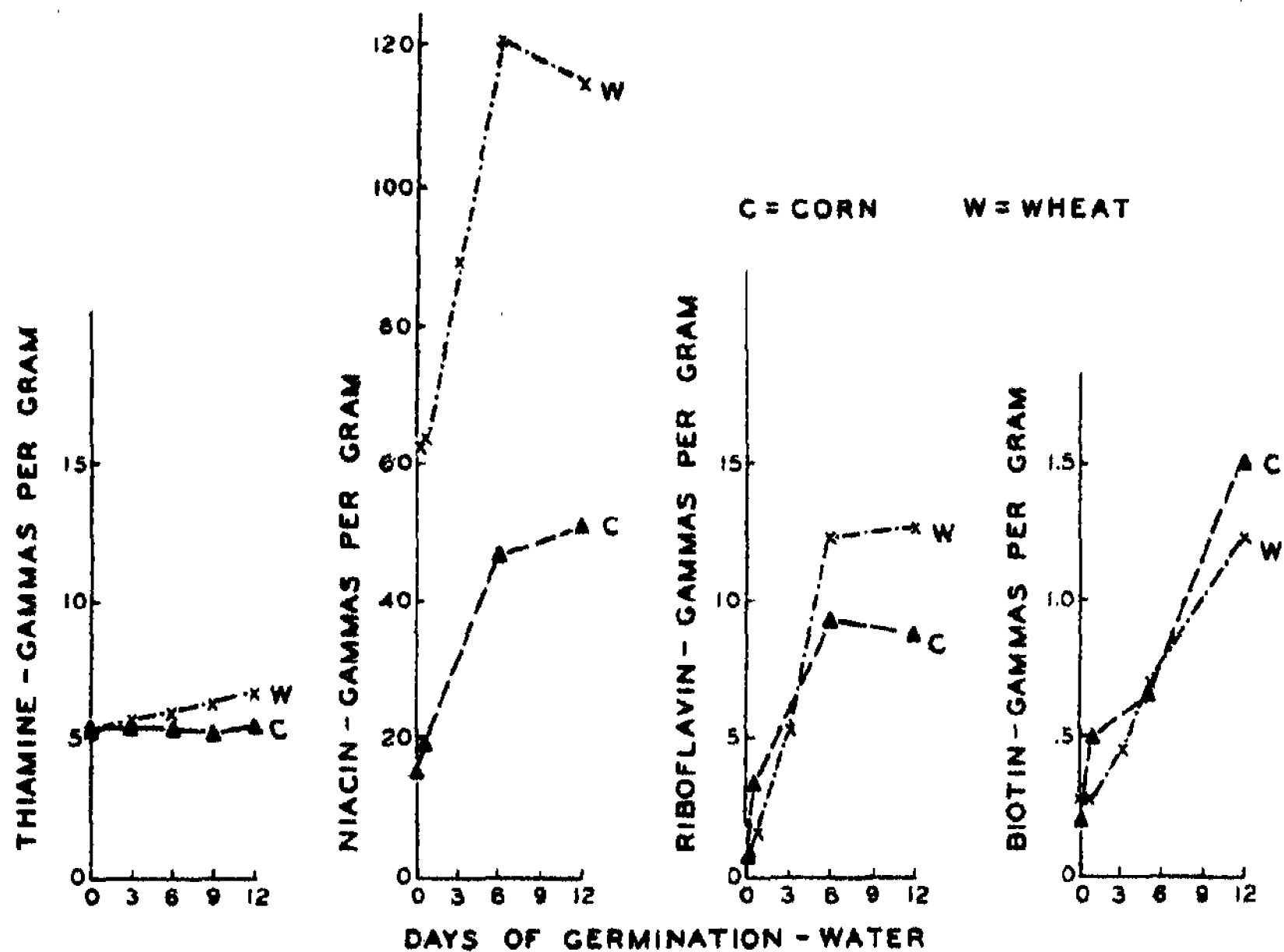


Figure 1 (above). Thiamine, niacin, riboflavin and biotin content of dry seeds and germinated seeds grown for different lengths of time in water cultures in the laboratory, July, 1942. Figure 2 (below). Vitamins in dry seeds and germinated seeds grown in sand cultures in a greenhouse, August, 1942. All data are expressed as micrograms (gamma) per gram of dry matter.

in the graph. In this experiment again, the thiamine content showed little change. Still other tests were performed on cereal seedlings grown in sand culture in the greenhouse and the resulting data, as shown in figure 2, further illustrate the characteristic upward trend of these three vitamins during germination. No significant changes in thiamine were observed in the corn and oat seedlings, but appreciable increases were noted in the wheat and barley seedlings. Pyridoxine appeared to reach much higher levels in sprouting barley and wheat than in germinating corn and oats, although marked increases were indicated in all the species of developing seedlings which were studied.

Determinations made on green leaves of young wheat, barley, corn and oats grown outdoors in soil indicated the presence of rich stores of B vitamins in these materials. The thiamine and riboflavin contents of young barley and wheat leaves were particularly striking in comparison with the amounts present in the dry grains. On a dry weight basis there was more than double the amount of thiamine and about 20 times more riboflavin present in the young green leaves than in the dry seed.

*Conclusions.*—Our investigations show significant increases in the concentration of riboflavin, niacin, biotin and pyridoxine during the germination of many kinds of edible seeds. In general, little or no change in thiamine concentration occurs during germination, though the content of green seedling leaves of cereals was found to be relatively high as compared with that of the dry grains. It should be stressed, furthermore, that the increased stores of B vitamins can be largely retained during the process of dehydrating the sprouted materials. If the value of germinated seeds is to be judged by their vitamin content, it appears that the common use of sprouted seeds in the diets of oriental peoples rests on a sound nutritional basis. It is hoped that this brief report may have some significance in connection with rational utilization of natural plant products in human and animal nutrition.

\* Grateful acknowledgment is made to the Nutrition Foundation for financial assistance given to Dr. George R. Cowgill and associates in botany.

<sup>1</sup> American Medical Association Monograph, *The Vitamins* (1939).

<sup>2</sup> Rosenberg, H. R., *Chemistry and Physiology of the Vitamins* (1942).

<sup>3</sup> Miller, Carey D., and Hair, Doris B., *Jour. Home Econ.*, 20, 263 (1928).

<sup>4</sup> Rytz, W., *Ber. Schweiz. Bot. Ges.*, 49, 339 (1939).

<sup>5</sup> Bonner, J., and Greene, J., *Bot. Gaz.*, 100, 226 (1938).

<sup>6</sup> Rose, Mary S., and Phipard, Esther H. F., *Jour. Nutrition*, 14, 55 (1937).

<sup>7</sup> Snell, E. E., and Strong, F. M., *Ind. Eng. Chem., Anal. Ed.*, 11, 346 (1939).

<sup>8</sup> Snell, E. E., Eakin, R. E., and Williams, R. J., *Jour. Am. Chem. Soc.*, 62, 175 (1940).

<sup>9</sup> Snell, E. E., and Wright, L. D., *Jour. Biol. Chem.*, 139, 675 (1941).

<sup>10</sup> Bonner, J., and Erickson, J., *Amer. Jour. Bot.*, 25, 685 (1938).

<sup>11</sup> Williams, R. J., Eakin, R. E., and McMahan, J. R., *University of Texas Publication*, No. 4137 (1941).

<sup>13</sup> Andrews, J. S., Boyd, H. M., and Terry, D. E., *Ind. Eng. Chem., Anal. Ed.*, 14, 271 (1942).

<sup>14</sup> Wegner, M. I., Kemmerer, A. R., and Fraps, G. S., *Jour. Biol. Chem.*, 144, 731 (1942).

---

## NOTES ON ALGAL NOMENCLATURE: I. POLLEXFENIA, JEANNERETTIA AND MESOTREMA

BY GEORGE F. PAPENFUSS

DEPARTMENT OF BOTANY, UNIVERSITY OF CALIFORNIA

Communicated September 1, 1942

Work on the marine algae of South Africa, carried on by the writer since 1935, has revealed a number of nomenclatural confusions. One of the most outstanding complexities centers around the usage of the generic name *Pollexfenia* Harvey.

When describing *Pollexfenia*, Harvey (1844, p. 431) stated: "This genus, founded on a plant from the Cape of Good Hope, is inscribed to the *Rev. John H. Pollexfen*. . . ." Harvey referred two species, both new, to *Pollexfenia*: *P. pedicellata* from Tasmania and *P. laciniata* from South Africa. Although giving *P. laciniata* second place, it is clear that Harvey intended it to be the nomenclatural type of his new genus.

The next reference to *Pollexfenia* in Harvey's publications is to be found in his *Nereis Australis* (1847). On page 22 he states: "This genus, named in honour of the *Rev. J. H. Pollexfen*, a successful explorer of the marine-botany of the Orkney Islands, contains two groups, which may hereafter be separated; perhaps they ought never to have been combined." Harvey accordingly split *Pollexfenia* into two subgenera: Subgenus 1, *Pollexfenia* in which he retained *P. pedicellata*; and subgenus 2, *Rhodoseris* in which he placed *P. (Rhodoseris) laciniata* and in addition a new species from Australia, *P. (Rhodoseris) cartilaginea* Harvey et Greville. In his *Index Generum Algarum* (1860, p. 5), Harvey raised the subgenus *Rhodoseris* to the rank of genus and designated as its type (by implication) *Pollexfenia laciniata*. This species thus came to serve as the type of both *Pollexfenia* and *Rhodoseris*.

In 1863 Harvey (1863, p. XVII) also transferred *Pollexfenia cartilaginea* to *Rhodoseris*. This species was subsequently designated as the type of the genus *Rhodoseris* by Schmitz (1889, p. 445) and Kylin (1924, p. 97) but these authors apparently overlooked the fact that Harvey in 1860 had cited *P. laciniata* as the type.

*Pollexfenia (Rhodoseris) laciniata* has been a little-known plant for a long

time, but has been collected at a number of localities in South Africa during recent years, both by Professor T. A. Stephenson and his co-workers and the writer. Not realizing that he had *P. laciniata* in hand, Kylin in 1938 redescribed the species under the name *Papenfussia elegans*, the type species of the genus *Papenfussia*.

From the preceding account it follows that the genus *Pollexfenia*, with the type *P. laciniata*, as originally proposed by Harvey should be retained for plants belonging to the Delesseriaceae; the genera *Rhodoseris* and *Papenfussia* become synonyms of *Pollexfenia*; and the Rhodomelaceae currently assigned to *Pollexfenia* have to be grouped under a different generic name. The name *Jeannerettia* Hook. et Harv. (in Harvey, 1847, p. 20), with the type *J. lobata* Hook. et Harv., is available for these species.

The plant that Harvey and Greville (in Harvey, 1847, p. 23) described as *Pollexfenia* (*Rhodoseris*) *cartilaginea* forms its tetrasporangia in special "leaflets" and accordingly is to be excluded from *Pollexfenia*, in which genus the reproductive organs are formed directly on the main thallus. *P. cartilaginea* may be representative of a new genus, but our knowledge concerning it is very meager and at present it seems best to refer the species to *Botryoglossum*, in which genus the reproductive organs are likewise formed in special "leaflets."

The following changes in nomenclature consequently seem warranted:

#### ***Pollexfenia* Harvey**

*London Jour. Bot.*, 3: 431, 1844. *Pollexfenia* subgenus *Rhodoseris* Harvey, *Ner. Austr.*: 22, 1847. *Rhodoseris* Harvey, *Index Gen. Alg.*: 5, 1860. *Papenfussia* Kylin, *Lunds Univ. Årsskr.*, N. F., Avd. 2, 34 (8): 15, 1938.

#### ***Pollexfenia laciniata* Harvey**

*London Jour. Bot.*, 3: 432, 1844. *Pollexfenia* (*Rhodoseris*) *laciniata* Harvey, *Ner. Austr.*: 22, pl. 6, 1847. *Rhodoseris laciniata* (Harv.) Harvey, *Index Gen. Alg.*: 5, 1860. *Papenfussia elegans* Kylin, *Lunds Univ. Årsskr.*, N. F., Avd. 2, 34 (8): 16, fig. 8, pl. 5, fig. 14, 1938.

#### ***Pollexfenia minuta* (Kylin) comb. nov.**

*Papenfussia minuta* Kylin, *Lunds Univ. Årsskr.*, N. F., Avd. 2, 34 (8): 17, pl. 5, fig. 15, 1938.

#### ***Botryoglossum cartilagineum* (Harv. et Grev.) comb. nov.**

*Pollexfenia* (*Rhodoseris*) *cartilaginea* Harv. et Grev. in Harvey, *Ner. Austr.*: 23, 1847. *Rhodoseris cartilaginea* (Harv. et Grev.) Harvey, *Phyc. Austr.*, 5: XVII, 1863.

**Jeannerettia** Hooker et Harvey<sup>1</sup>

in Harvey, *Ner. Austr.*: 20, 1847. *Pollexfenia* subgenus *Pollexfenia* Harvey, *Ner. Austr.*: 22, 1847. *Pollexfenia* Harvey, *Phyc. Austr.*, 5: XVII, 1863; Schmitz und Falkenberg in Engler und Prantl, *Nat. Pflanzenfam.*, 1 (2): 454, 1897; Falkenberg, *Rhodomelaceen*: 290, 1901. *Melanoseris* Zanardini, *Flora*, 57: 489, 1874.

**Jeannerettia lobata** Hooker et Harvey

in Harvey, *Ner. Austr.*: 20, pl. 4, 1847. *Pollexfenia lobata* (Hook. et Harv.) Falkenberg, *Rhodomelaceen*: 295, 1901.

**Jeannerettia pedicellata** (Harv.) comb. nov.

*Pollexfenia pedicellata* Harvey, *London Jour. Bot.*, 3: 431, 1844.

**Jeannerettia crispata** (Zanard.) comb. nov.

*Melanoseris crispata* Zanardini, *Flora*, 57: 489, 1874. *Pollexfenia crispata* (Zanard.) Falkenberg in Engler und Prantl, *Nat. Pflanzenfam.*, 1 (2): 455, fig. 256, 1897.

**Jeannerettia nana** (J. Ag.) comb. nov.

*Pollexfenia nana* J. Agardh, *Anal. Alg.*: 164, 1892.

**Jeannerettia crenata** (J. Ag.) comb. nov.

*Pollexfenia crenata* J. Agardh, *Anal. Alg.*: 165, 1892.

In a recent paper it was pointed out by J. De Toni (1936) that *Martensia* Hering (1841), founded on *M. elegans* Hering from South Africa, is invalidated by *Martensia* Giseke (1792), a genus of flowering plants, and he accordingly proposed the generic name *Capraella* for the species of red algae currently grouped under *Martensia* Hering. De Toni apparently overlooked the fact that the generic name *Mesotrema* J. Agardh (1854) is available for these plants. This genus was erected by J. Agardh to receive a new species, *Mesotrema Pavonia* J. Ag., from the West Indies. When describing *Mesotrema*, J. Agardh was aware of its close relationship with *Martensia* Hering and in 1863 he immersed it, species and all, in the latter genus. As originally proposed and described, however, *Mesotrema* seems to conform to the International Rules of Botanical Nomenclature (1935) and should be retained. *Hemitrema* R. Brown (1843), with the type *H. Kraussii* R. Brown, is based on *Martensia elegans* Hering (1841) and is invalid.

**Mesotrema** J. Agardh

*Öfvers. Kongl. Sv. Vetensk.—Akad. Förhandl.*, 11: 110, 1854. *Martensia*

Hering, *Ann. and Mag. Nat. Hist.*, **8**: 92, 1841 (non *Martensia* Giseke, 1792). *Hemitrema* R. Brown in Endlicher, *Mant. Bot. sistens Gen. Plant.*, Suppl. 3: 50, 1843. *Capraella* J. De Toni, *Not. nomencl. alg.*, 7, 1936.

**Mesotrema Pavonia** J. Agardh

*Öfvers. Kongl. Sv. Vetensk.—Akad. Förhandl.*, **11**: 110, 1854. *Martensia Pavonia* (J. Ag.) J. Agardh, *Sp. Alg.*, **2** (3): 831, 1863. *Capraella Pavonia* (J. Ag.) J. De Toni, *Not. nomencl. alg.*, 7, 1936.

**Mesotrema elegans** (Hering) comb. nov.

*Martensia elegans* Hering, *Ann. and Mag. Nat. Hist.*, **8**: 92, 1841. *Capraella elegans* (Hering) J. De Toni, *Not. nomencl. alg.*, 7, 1936. *Hemitrema Kraussii* R. Brown in Endlicher, *Mant. Bot. sistens Gen. Plant.*, Suppl. 3: 50, 1843.

**Mesotrema fragilis** (Harv.) comb. nov.

*Martensia fragilis* Harvey, *Hooker's Jour. Bot.*, **6**: 145, 1854. *Capraella fragilis* (Harv.) J. De Toni, *Not. nomencl. alg.*, 7, 1936.

**Mesotrema denticulata** (Harv.) comb. nov.

*Martensia denticulata* Harvey, *Trans. Royal Irish Acad.*, **22**: 537, 1855. *Capraella denticulata* (Harv.) J. De Toni, *Not. nomencl. alg.*, 7, 1936.

**Mesotrema australis** (Harv.) comb. nov.

*Martensia australis* Harvey, *Trans. Royal Irish Acad.*, **22**: 537, 1855. *Capraella australis* (Harv.) J. De Toni, *Not. nomencl. alg.*, 7, 1936.

**Mesotrema flabelliformis** (Harv. ex J. Ag.) comb. nov.

*Martensia flabelliformis* Harvey ex J. Agardh, *Sp. Alg.*, **2** (3): 826, 1863. *Capraella flabelliformis* (Harv.) J. De Toni, *Not. nomencl. alg.*, 7, 1936.

**Mesotrema speciosa** (Zanard.) comb. nov.

*Martensia speciosa* Zanardini, *Flora*, **57**: 488, 1874. *Capraella speciosa* (Zanard.) J. De Toni, *Not. nomencl. alg.*, 7, 1936.

**Mesotrema Beccariana** (Zanard.) comb. nov.

*Martensia Beccariana* Zanardini, *Nuovo Giorn. Bot. Ital.*, **10**: 35, 1878.

The writer wishes to express his appreciation to Professor A. R. Davis for granting the facilities of the Botanical Laboratories of the University

of California and to Professor Emeritus William A. Setchell for the use of his excellent library. The study was begun with the assistance of a grant from the Carnegie Corporation of New York through the University of Cape Town and was completed during the tenure of a Fellowship granted by the Carnegie Corporation.

<sup>1</sup> The systematic position of *Pollexfenia tenella* Kützinger (1849, p. 875) is obscure. De Toni (1903, p. 981) retains it, with a query, in *Pollexfenia* (*Jeannerettia* in the sense of the present article), but suspects relationship with *Symphocladia marchantioides*, under which species he has also included it as a doubtful synonym. According to Cotton (1915, p. 202), the type specimen is not to be found in Herbarium Kützinger.

Agardh, J. G.

1854. "Nya algformer," *Öfvers. Kongl. Sv. Vetensk.—Akad. Förhandl.*, 11: 107–111.

1863. *Species, genera et ordines algarum*, 2 (3). Lundae.

1892. "Analecta Algologica," *Actis Soc. Physiogr. Lundensis*, 28: 1–182.

Brown, R.

1843. *Hemitrema* in S. L. Endlicher, *Mant. Bot. sistens Gen. Plant.*, Suppl. 3: 50. Vindobonae.

Cotton, A. D.

1915. "Cryptograms from the Falkland Islands Collected by Mrs. Vallentin," *Jour. Linn. Soc., Botany*, 43: 137–231.

De Toni, J. B.

1903. *Sylloge Algarum*, 4 (3). Patavii.

De Toni, J.

1936. *Noterelle di nomenclature algologica. VII. Primo elenco di Floridee omonime.* Published by the author. Brescia.

Engler, A., und Prantl, K.

1897. *Die natürlich. Pflanzenfam.*, 1 (2). Leipzig.

Falkenberg, P.

1901. "Die Rhodomelaceen des Golfes von Neapel und der angrenzenden Meeres-Abschnitte." *Fauna und Flora des Golfes von Neapel*, Monogr. 26.

Harvey, W. H.

1844. "Algae of Tasmania," *London Jour. Bot.*, 3: 423–454.

1847. *Nereis Australis*. London.

1854. "Short Characters of Three New Algae from the Shores of Ceylon," *Hooker's Jour. Bot.*, 6: 143–145.

1855. "Some Account of the Marine Botany of the Colony of Western Australia," *Trans. Royal Irish Acad.*, 22: 525–566.

1860. *Index Generum Algarum*. London.

1863. *Phycologia Australica*, Vol. V. London.

Hering, D.

1841. "Diagnoses algarum novarum a cl. Dre. Ferdinand Krauss in Africa australi lectarum," *Ann. and Mag. Nat. Hist.*, 8: 90–92.

Kützinger, F. T.

1849. *Species Algarum*. Lipsiae.

Kylin, H.

1924. "Studien über die Delesseriaceen," *Lunds Univ. Årsskr., N. F., Avd. 2*, 20 (6).

1938. "Verzeichnis einiger Rhodophyceen von Südafrika," *Ibid.*, 34 (8).

Schmitz, F.

1889. "Systematische Übersicht der bisher bekannten Gattungen der Florideen," *Flora*, 72: 435-456.

Zanardini, J.

1874. "Phyceae Australicae novae vel minus cognitae," *Flora*, 57: 488-490, 497-505.  
1878. "Phyceae Papuanae novae vel minus cognitae a cl. O. Beccari in itinere ad Novam Guineam annis 1872-75 collectae," *Nuovo Giorn. Bot. Ital.*, 10: 34-40.





# PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

Volume 28

November 15, 1942

Number 11

Copyright 1942 by the National Academy of Sciences

## *SOME FACTORS AFFECTING THE TOXICITY OF CULTURES OF SHIGELLA DYSENTERIAE*

BY RENÉ J. DUBOS,\* HENRY D. HOBERMAN† AND CYNTHIA PIERCE

HARVARD SCHOOLS OF MEDICINE AND PUBLIC HEALTH

Communicated October 15, 1942

Cultures of *Shigella dysenteriae* are extremely toxic to man and experimental animals; it is likely that this toxicity is of great significance in determining the symptomatology and pathology of Shiga dysentery. Many conflicting views have been voiced to account for it. Some authors, for instance, feel that all the toxic manifestations are due to one single substance.<sup>1</sup> Others, on the contrary, believe in the existence of two independent toxins<sup>2-8</sup>: (a) an endotoxin, which exhibits a special affinity for the intestinal tract, is present only in the smooth variants of the organism, is associated chemically with the specific somatic polysaccharide and is resistant to heat and to proteolytic enzymes; (b) an exotoxin with neurotropic affinity, which is present both in the rough and smooth variants, is of protein nature and is inactivated by heat and trypsin. It has also been claimed that aerobic incubation favors the production of the exo-neurotoxin whereas only the endotoxin is produced under anaerobic conditions.<sup>2</sup>

It is obvious that, before any final claim can be made concerning the nature and pathological activity of the toxic components of the Shiga bacillus, methods have to be devised for the production in large quantities and for the purification of these substances. In view of the practical importance of this problem for the eventual development of immunization procedures, we are reporting at the present time some of our preliminary observations which may serve to define better the optimal conditions for the production of toxin and the properties of at least one of the toxic fractions.

*Effect of Cultural Conditions on the Toxicity of Cultures of Shigella dysenteriae.*—All experiments were performed with strain 32158 of *Shigella dysenteriae* which was obtained through the courtesy of Miss M. Coleman of the New York State Department of Health. Two phase variants were used: (1) the smooth form (s) which gave typical agglutination in standard antisera, and (2) a rough variant ( $R_{12}$ ) which was isolated from an old stock

culture. The organisms grown in liquid media were collected by centrifugation or filtration (in the case of the *R* variant); the agar cultures were suspended in distilled water and also centrifuged. It was observed that in the case of young cultures at least (24 to 48 hours old, for instance), only a very small percentage of the total toxicity of the whole culture was lost in the culture supernatant or cell washings; in other words, the toxic components of young cultures seemed to be bound to the cell structure, a fact already observed by others.<sup>9</sup>

The dry weights of the cultures were obtained by desiccation over phosphorus pentoxide *in vacuo*, or by treatment with acetone ether; it was found that these methods of desiccation did not alter the toxicity of the original material.

A number of different toxicity tests were used: minimal lethal dose for rabbits and for white mice (20 grams), production of the Shwartzman reaction in rabbits,<sup>10</sup> production of the ocular reaction in rabbits,<sup>11</sup> etc. Rabbits were found to be much more susceptible than mice; for reasons of economy, however, most titrations of toxicity were based on determination of the minimum lethal dose for mice, and only the results of these tests will be reported at the present time.

The effect of cultural conditions on toxicity is well illustrated by the following experiment. Cultures of rough and smooth variants of Shiga bacillus were grown at 37° for 18 hours in meat infusion peptone broth, (the layer of fluid being 10 cm. thick), or on meat infusion peptone agar. The washed cells were desiccated *in vacuo* over phosphorus pentoxide, re-suspended in 0.05 *M* phosphate buffer at pH 7.2, heated at 60°C. for 10 minutes, and injected intraperitoneally into white mice. The results presented in table 1-A indicate that in the case of both the rough and smooth variants, 0.15 mg. was the minimal lethal dose when the agar cultures were used, and 1.5–2.0 mg. with the broth cultures.

Aqueous extracts of agar were added to meat infusion peptone broth to determine whether there is present in agar a water-soluble component which favors toxin production; no increase in toxicity of the broth culture, however, resulted from this treatment.

It appeared possible that the greater production of toxin on agar media was due to the fact that the agar gel permits surface growth and therefore increases aerobic metabolism. To test this hypothesis, media were prepared in which the solid surface was supplied by silica gel<sup>12</sup> instead of agar. To the silica gel were added 0.05 *M* phosphate buffer at pH 7.4, 1.0 per cent peptone and 1.0 per cent meat extract. The medium was poured into petri dishes, some of the plates receiving 1.0 per cent dextrose in addition to the other components of the medium. The plates were inoculated with rough or smooth variants of Shiga cultures; after 18 hours' incubation, the growth (which was less abundant than on agar) was collected from the

TABLE 1  
TOXICITY OF DESICCATED CELLS OF *R* AND *S* *Shigella dysenteriae* GROWN UNDER DIFFERENT CONDITIONS

CULTURE		MEDIUM	NUMBER OF MICE SURVIVING I.P. INJECTION OF FOLLOWING AMOUNTS OF CELLS (MG. DRY WEIGHT)			
			1-A			
			1.5	1.0	0.3	0.1
R	Meat infusion peptone agar		0/3 <sup>a</sup>	0/3	0/3	1/3
R	Meat infusion peptone broth		2/3	2/3	3/3	3/3
S	Meat infusion peptone agar		0/3	0/3	0/3	0/3
S	Meat infusion peptone broth		2/3	3/3	3/3	3/3
			1-B			
			1.0	0.5	0.25	0.1
R	Nutrient silica gel		0/1	0/2	0/3	0/2
R	Nutrient silica gel + 1% glucose		0/1	0/2	2/3	3/3
S	Nutrient silica gel		...	0/3	0/3	0/3
S	Nutrient silica gel + 1% glucose		...	0/3	2/3	3/3
			1-C			
			1.5	0.75	0.45	0.15
R	Peptone broth mechanically shaken (total yield 600 mg./liter)		0/3	0/3	1/3	2/3
R	Anaerobic peptone broth (total yield 200 mg./liter)		1/3	3/3	3/3	3/3
			1-D			
			1.5	0.75	0.30	0.15
R	Peptone broth mechanically shaken (total yield 750 mg./liter)		...	0/3	2/3	3/3
R	Thioglycollate peptone broth (total yield 420 mg./liter)		3/3	3/3	..	..

<sup>a</sup> The denominator indicates the number of mice used in the test. The numerator indicates the number of mice surviving after 10 days.

silica gel surface, washed with water, desiccated with acetone ether, resuspended in neutral phosphate buffer, heated at 60°C. for 10 minutes and injected intraperitoneally into white mice. The results presented in table 1-B indicate that, in the case of both the rough and smooth variants, the cells grown on silica gel medium not containing glucose were extremely toxic, even more toxic than the cultures grown on agar media. Addition of glucose to the medium markedly reduced toxicity.

Since surface growth gives rise to cells much more toxic than the cells grown in liquid media, an attempt was made to increase oxidative conditions within the broth by bubbling oxygen during the whole course of incubation; these conditions brought about some increase in toxicity, but the

results were not striking enough to warrant further description at the present time.

In other experiments, the inoculated medium was violently agitated on a shaking machine during the whole period of growth. It was found that, not only did the shaking markedly increase the total yield of cells, but that the toxicity per milligram of dry weight of cell was also greatly increased. These facts are illustrated in the two following experiments.

A medium containing 0.05 *M* phosphate at pH 6.9, 1.0 per cent peptone (Difco tryptone) and 1.0 per cent glucose was inoculated with a rough variant of *Shigella dysenteriae*. Half of the medium was incubated in an aerobic jar;<sup>18</sup> the other half was violently agitated on a shaking machine during the whole incubation period. The cells were collected by centrifugation after acidification at pH 4.5 (a reaction at which none of the broth constituents are precipitated), washed with water and desiccated with acetone ether. The total yields were approximately 600 mg. per liter in the case of the aerobic culture, and only 200 mg. in the case of the anaerobic culture. The cells were resuspended in neutral phosphate buffer, heated at 60°C. for 15 minutes and injected intraperitoneally into white mice for a determination of comparative toxicity (table 1-C).

In another experiment anaerobic conditions of incubation were obtained by the addition of 0.05 per cent of sodium thioglycollate to the medium, and aerobic conditions secured by constant agitation. The medium consisted of 1.0 per cent tryptone (Difco), 1.0 per cent glucose, 0.05 *M* phosphate at pH 7.8, and was inoculated with a rough variant of Shiga bacillus (table 1-D).

The results presented in tables 1-C and 1-D indicate clearly that anaerobic incubation gives yields of cells very much smaller than those obtained under conditions of aerobic agitation during growth (750 mg. per liter versus 420). The toxicity of the cells grown under aerobic conditions is also much greater per milligram dry weight than that of cells obtained by anaerobic incubation.

*Separation of a Toxic Component from Cultures of a Rough Variant of Shigella dysenteriae.*—Large amounts of culture of the variant form *R*<sub>12</sub> of the *Shigella dysenteriae* were grown on meat infusion peptone agar at 37°C. for 18 hours; the cells were washed in water and desiccated *in vacuo* over phosphorus pentoxide. The average yield was approximately 20 mg. of washed dried cells per agar plate (8 cm. diameter); 0.15 mg. of dried cells (heated at 60°C. for 15 minutes in 0.05 *M* neutral phosphate buffer) was the minimal lethal dose for mice. The bacterial component responsible for the toxicity of the culture exhibits the following properties: (1) it is only slowly inactivated by heat at neutral and at acid reactions, but is very rapidly destroyed at alkaline pH; (2) it is completely resistant to proteolytic enzymes (crude trypsin, pepsin, papain, mold and bacterial en-

zymes); (3) it is precipitated by two-thirds saturation with ammonium sulfate, and by 33 to 50 per cent acetone; (4) it is precipitated at acid pH (pH 3.0 to 4.0); (5) it can be obtained in solution in 0.05 *M* dibasic phosphate from heat-killed cultures digested with trypsin; (6) the addition of papain to a solution of the toxic material causes the precipitation of an inactive fraction (probably nucleic acid), whereas the active material remains in solution; it can be precipitated again from the papain solution by addition of acid; (7) the active principle does not dialyze through cellophane membranes.

By making use of these different properties, we have obtained fractions of such activity that 3.0 micrograms injected intraperitoneally into white mice, or intravenously into rabbits, regularly causes death within 3 to 4 days; death of mice was also observed with amounts as small as 1.3 microgram. The material is stable and retains its activity unaltered for several weeks when kept in neutral solution at 5°C.

*Conclusions.*—It has been shown that conditions which favor aerobic metabolism (growth on agar or on silica gels, or in broth violently agitated during incubation) greatly increase the yield of *S. dysenteriae* cultures and the toxicity of the cells (measured in terms of dry weight). For instance, 1.5 mg. of dried cells grown in an anaerobic jar failed to kill any mice, whereas 0.1 mg. of cells grown on silica gel plates killed all mice (smaller amounts were not tried). The addition of glucose to silica gel medium caused a definite decrease in the toxicity of the cells; this may be due to the fact that, even when growth takes place under aerobic conditions, the presence of glucose permits some anaerobic metabolism.

The effect of environmental conditions of growth on the yield of organisms and on the yield of toxin was the same for rough and for smooth variants of *Shigella dysenteriae*. In all cases also, the largest percentage of total toxicity of the whole culture was found to be bound to the bacteria bodies and to be released in solution only as a result of autolysis.

The toxic factor is resistant to proteolytic enzymes; it can be obtained in the form of a water-soluble fraction which is stable and of which 1.0 to 3.0 micrograms cause the death of mice and rabbits within 72 hours.

\* Preliminary experiments were carried out in the laboratories of the Hospital of the Rockefeller Institute, New York.

† Fellow in Medical Sciences of the National Research Council.

<sup>1</sup> Waaler, E., *Jour. Exper. Med.*, **63**, 1-16 (1936).

<sup>2</sup> Olitzki, L., and Leibowitz, J., *Brit. Jour. Exper. Path.*, **16**, 523-531 (1935).

<sup>3</sup> Boivin, A., and Mesrobian, L., *Compt. Rend. Soc. Biol.*, **124**, 439, 442, 1078, 1092 (1937).

<sup>4</sup> Boivin, A., *Ibid.*, **133**, 252-255 (1940).

<sup>5</sup> Haas, R., *Ztschr. Immunitäts. exper. Therap.*, **91**, 254-272 (1937).

<sup>6</sup> Haas, R., *Ibid.*, **99**, 121 (1940).

- <sup>7</sup> Istrati, Gh., *Compt. Rend. Soc. Biol.*, **129**, 1010-1013 (1938).  
<sup>8</sup> Wagner-Jauregg, Th., *Centralbl. f. Bakt.*, **144**, I. Orig., 31-32 (1939).  
<sup>9</sup> Okell, C. C., and Blake, A. V., *Jour. Path. & Bact.*, **33**, 57-63 (1930).  
<sup>10</sup> Shwartzman, G., "Phenomenon of Local Tissue Reactivity," N. Y., 461 pp. (1937).  
<sup>11</sup> Ajo, C., *Proc. Soc. Exper. Biol. & Med.*, **47**, 500-501 (1941).  
<sup>12</sup> Sterges, A. J., *Jour. Bact.*, **43**, 317-327 (1942); *Ibid.*, **44**, 138 (1942).  
<sup>13</sup> Mueller, J. H., and Miller, P. A., *Ibid.*, **41**, 301-303 (1941).

## THE FUSION OF BROKEN ENDS OF CHROMOSOMES FOLLOWING NUCLEAR FUSION

BY BARBARA McCLINTOCK

DEPARTMENT OF GENETICS, CARNEGIE INSTITUTION OF WASHINGTON, COLD SPRING HARBOR

Communicated September 22, 1942

When, through radiation or other causes, chromosomes are broken within a single nucleus, 2-by-2 fusions may occur between the broken ends. These fusions may lead to rearrangements of parts of the chromatin complement, giving rise to various chromosomal aberrations which are detected as reciprocal translocations, inversions, deficiencies, etc. Since, in the well-investigated cases, the breakages occurred within a single nucleus, the conditions that lead to fusions of broken ends could not easily be ascertained. The following questions have been asked: (1) Must two or more chromosomes be in intimate contact at the time of breakage in order that fusions may occur? (2) If no intimate contact is necessary at the time of breakage, are the broken ends "unsaturated," that is, capable of fusion with any other unsaturated broken end? (3) If question (2) can be answered in the affirmative, what forces are involved which lead to the contact and subsequent fusions of the two unsaturated broken ends? Likewise, (4) how long will these broken ends remain unsaturated, i.e., capable of fusion?

Questions (1) and (2) could be answered if the following conditions were present: Assume that fusion occurs between two nuclei each of which possesses one chromosome, one end of which has been broken. Each nucleus will then have a single broken end. When these nuclei fuse and their chromosomes intermix within a single nucleus, the chromosome with a broken end contributed by one nucleus could fuse with the chromosome with a broken end contributed by the second nucleus. The chromosome fusion should occur between these two broken ends. This experiment may easily be conducted in maize. The two nuclei that fuse can be the male and the female gametes, respectively. The method of obtaining

gametes having a chromosome with a single "unsaturated" broken end has been reported previously.<sup>1</sup> This method may be briefly summarized. Plants were obtained which possessed one normal chromosome 9 and one chromosome 9 with a duplication of the short arm. This duplicated arm extended beyond the normal short arm, the serial order of parts within the duplicated segment being the reverse of that of the normal short arm. When the duplicated segment is involved in crossing-over, a dicentric chromatid may be produced which is the equivalent of two chromosomes 9 attached at the ends of their short arms. Breakage of this dicentric chromatid during a meiotic anaphase results in the entry into a spore nucleus of a chromatid with a single broken end. Fusion then occurs at the position of breakage between the two sister halves of this broken chromatid, forming a new dicentric chromatid. This, in turn, produces a bridge configuration in the first gametophytic division, as the two centromeres of the dicentric chromatid pass to opposite poles in the anaphase figure. Again, a broken chromatid enters each telophase nucleus. In each nucleus, fusion again occurs between the two sister halves of this broken chromatid at the position of this latter breakage. This *chromatid* type of breakage-fusion-bridge cycle continues in successive gametophytic mitoses. Therefore, all the nuclei of the fully developed male or female gametophyte will possess one chromosome with a single broken end. Following fertilization, two nuclei from the female gametophyte and one from the male gametophyte fuse to form the primary endosperm nucleus. If the nuclei of one of the gametophytes possesses such a chromosome carrying dominant genes in the arm with the broken end, and if the other gametophyte possesses a normal, non-broken chromosome carrying the recessive alleles, variegation for these genes will be apparent in the fully developed endosperm. This is because the chromosomes with the broken ends continue the breakage-fusion-bridge cycle in successive nuclear divisions during the development of the endosperm. Following various non-median breakages of the bridge configurations, the dominant genes may be deleted from some nuclei and duplicated in the sister nuclei. Continued repetition of this type of breakage during the development of the endosperm produces a conspicuous variegation pattern. The behavior of the chromosome with the broken end in the sporophytic tissues of these kernels is entirely different. The chromatid type of breakage-fusion-bridge cycle ceases when the zygote is formed. The newly broken end "heals." Following this healing, no further fusions occur. The broken end is as stable in its subsequent behavior as any normal chromosome end.

There is no reason to suspect that the condition of those chromosomes in the gamete nuclei that participate in endosperm fusions differ from those participating in zygotic fusions. However, in one case, the broken end remains unsaturated, and in the other case the broken end heals. It is



reasonable to believe, therefore, that the healing process occurs subsequent to zygotic fusions and not before. One may tentatively assume that a chromosome end, broken in the pre-gametic division, is unsaturated at the time of zygotic fusion. If each gamete contributes a chromosome with an unsaturated broken end, one could expect fusion to occur between these two broken ends. The reasoning behind this expectation is based on the behavior of ring-shaped chromosomes in sporophytic tissues.<sup>2</sup> Ring-shaped chromosomes are frequently broken during mitotic anaphases. Following such breakage, the telophase nuclei receive a chromosome both ends of which are broken. Fusion may then occur between these unsaturated broken ends, reestablishing the ring-shape. This is a chromosome fusion, not a chromatid fusion. This behavior has suggested that fusion of unsaturated broken ends in sporophytic tissues may be chromosomal in contrast to gametophytic and endosperm tissues where chromatid fusions may occur. The experiment to be described furnishes proof of the correctness of these assumptions.

To test whether the broken ends are unsaturated in the gamete nuclei, two such ends were introduced into the zygote nucleus, one contributed by the male gamete and one by the female gamete. Detection of kernels whose zygote nuclei had received such chromosomes was accomplished by introducing contrasting endosperm markers carried by the chromosomes with the broken ends (*i*, aleurone color, and *wx*, waxy starch, located in the short arm of one parental chromosome 9 and the alleles *I*, inhibitor of aleurone color, and *Wx*, normal starch, carried by the other parental chromosome 9). The endosperms of those kernels that receive a broken chromosome 9 from each parent should show a very distinctive type of variegation. All three broken chromosomes 9 would undergo the break-fusion-bridge cycle. This would lead to variegation for *I-i* and *Wx-wx*, variegation for depth of color in the *i* regions due to multiplication of the number of *i* genes, and scarred and pitted regions in both the *I* and *i* sectors due to the presence of cells which are homozygous deficient for segments of the short arm of chromosome 9.<sup>3</sup>

Plants heterozygous for the duplication chromosome 9 and homozygous for *i wx* were crossed by plants heterozygous for the duplication and homozygous for *I* and *Wx*. Three types of kernels, with respect to endosperm characters, should be produced; (1) *I Wx*, non-variegated kernels following fusion of a nucleus carrying a non-broken *I Wx* chromosome with nuclei carrying either a non-broken or a broken *i wx* chromosome; (2) kernels variegated for *I-i* and *Wx-wx* following fusion of a nucleus carrying a broken *I Wx* chromosome with nuclei carrying a non-broken *i wx* chromosome; (3) kernels resulting from the fusion of a nucleus carrying a broken *I Wx* chromosome with nuclei carrying a broken *i wx* chromosome. As stated above, the endosperm character of this latter type of kernel could be

anticipated. When observations were made of the kernels resulting from this cross, these latter kernels were very conspicuous. From a total of 18,243 kernels examined, 20 possessing an embryo were of this latter type. More of this type were present but they were germless. These 20 kernels were germinated to determine what had happened to the two broken chromosomes 9 which had entered the zygote. If both broken ends had healed without fusion, normal-appearing plants would be expected to arise from these kernels. If the two broken ends had fused, a dicentric chromosome would have been produced. It would be composed of the chromosome 9 contributed by the male gamete and the chromosome 9 contributed by the female gamete, with their short arms fused end-to-end. When this dicentric chromosome divided and when the two centromeres of each chromatid passed to opposite poles at a mitotic anaphase, two contiguous bridges should be formed. Following breakage of these two bridges, two chromosomes, each with a freshly broken end, should enter each sister telophase nucleus. As stated previously, one could expect fusions to occur between the broken ends of these two chromosomes in each sister telophase nucleus. This would reestablish the dicentric condition, for again the two chromosomes 9 would be joined to form one chromosome with two centromeres. Repeated anaphase bridge configurations should be expected in subsequent divisions following this *chromosomal* type of breakage-fusion-bridge cycle. The cells of a plant possessing such a dicentric chromosome undergoing this behavior should be composed of various types of homozygous and heterozygous duplications and deficiencies of the short arm of chromosome 9, following repeated non-median breaks in the anaphase bridges. Consequently, these plants should be conspicuously modified in appearance, because of the variation in degree of duplication or deficiency in the many nuclei of the plant.

In the seedling stage, 10 of the 20 plants arising from the kernels classified as having received a broken chromosome 9 from each parent were obviously of the type expected if a dicentric chromosome were present. The presence of the dicentric chromosome was confirmed by examination of the division figures in the young roots, where nearly half of the observed anaphase figures showed two contiguous bridges derived from a dicentric chromosome. Nine of the remaining plants were normal in appearance, and one plant was normal in morphological growth but pale yellow<sup>4</sup> in color and died in the seedling stage. No bridges were observed in the roots of these latter 10 plants.

Due to aberrant growth and death of many cells, 5 of the plants with a dicentric chromosome died in the seedling stage. The remaining 5 plants continued to grow. In all 5 plants, as growth continued, sectors of tissue developed which showed no aberrant growth patterns. These sectors were quite normal in appearance. Gradually, these recovered sectors gained the

ascendency in growth until most of the plant was normal in appearance. In one plant, 3 normal side shoots developed from the base of a very aberrant main shoot which was obviously dying. Root tips were taken at various times from all of the 5 plants that had survived the seedling stage. In all cases, vigorous growth of some side branches of the root system was noted. Examination of division figures in these roots no longer showed any dicentric bridges. The examined cells possessed the normal chromosome number of 20, instead of the 18 monocentric chromosome plus 1 dicentric chromosome previously observed in the younger roots. Sporocytes for examination of the chromosomes at pachytene were obtained from two of the three recovered shoots of one plant and from the recovered main shoots of the four other plants. In all cases, 10 bivalent chromosomes were present, one of which was a bivalent chromosome 9. The two chromosomes 9 were not fused at the ends of their short arms. In most cases, the composition of the short arm of each member of the bivalent was greatly modified, although in each tassel sample the two chromosomes 9 maintained their respective morphologies in all examined cells. Of the 12 chromosomes 9 examined from these six samples, no two were alike. The composition of the short arms of the chromosomes 9 in the two recovered branches from one plant originally possessing a dicentric chromosome was entirely different. In each case, it was apparent that the cells of the examined part of the tassel had originated from one individual cell whose cell ancestors had previously been undergoing the chromosomal type of breakage-fusion-bridge cycle involving the original dicentric chromosome 9. This could be determined in each case by the comparative morphologies of the short arms of the two members of the bivalent. In several cases it was possible to determine the minimum number of fusions, bridges and breaks that must have occurred before "healing" of the two broken ends had taken place in the nucleus of the particular cell that gave rise to the recovered sector. The factors involved in the process of healing of two such broken ends within a single nucleus are still undetermined.

The pachytene chromosomes were examined in the surviving 9 of the 10 plants which were normal in morphological growth from the earliest seedling stage and which showed no bridges in the earliest roots. This examination showed that 4 of these plants had received a broken chromosome 9 from each parent. However, morphological analysis of the short arms of the two chromosomes 9 in each case gave no indication that fusions had ever occurred between their broken ends. In one plant, one parent had contributed a broken chromosome 9, but it was not possible to determine whether the other parent had contributed a broken chromosome 9. In the remaining 4 plants, each parent had contributed a broken chromosome 9. However, the broken end of one chromosome 9 had fused with a broken end of a chromosome other than that of the chromosome 9 introduced by the

second gamete. In each case, the broken end of the second chromosome 9 had no unsaturated end with which it could unite. As expected, this single broken end thereafter healed.

*Conclusions.*—The experiments outlined allow some specific answers to be given to the questions presented in the first paragraph of this paper. Question (1) may be answered in the negative. Two chromosomes do not have to be in contact at the time of breakage in order that fusions may occur between their broken ends. This was shown by the fusion that occurred in the zygote or in an early embryonic nucleus between a broken end of chromosome 9 contributed by the male gamete and a broken end of the chromosome 9 contributed by the female gamete. Question (2) may be answered in the affirmative. This was shown by the fusion of these two chromosomes, which produced a dicentric chromosome, and the subsequent behavior of this dicentric chromosome which, for some time, followed the chromosomal type of breakage-fusion-bridge cycle. Question (3) cannot be answered directly from the present observations. Nevertheless, the observations imply that some force exists which accounts for the fusion of unsaturated broken ends of chromosomes. Question (4) likewise cannot be answered directly. Nevertheless, it is certain that the unsaturated state does not persist indefinitely. An unsaturated broken end will become saturated or healed and incapable of fusions when only one such broken end is present in sporophytic tissues;<sup>1</sup> or, as shown in this report, two such broken ends may heal without fusions even when these two ends are present in the same nucleus.

<sup>1</sup> McClintock, B., *Genetics*, **26**, 234–282 (1941).

<sup>2</sup> McClintock, B., *Ibid.*, **23**, 315–376 (1938).

<sup>3</sup> McClintock, B. (unpublished).

<sup>4</sup> This mutant type appears when a plant is homozygous deficient for a small terminal segment of the short arm of chromosome 9.

---

## NON-RANDOM UNCOILING IN HETEROBRACHIAL CHROMOSOMES

BY A. H. SPARROW\*

BIOLOGICAL LABORATORIES, HARVARD UNIVERSITY

Communicated October 2, 1942

In theories regarding the mechanics of chromosome spiralization much depends upon whether there is a definite (inherent?) pattern of coiling or whether the direction of coiling is purely a matter of chance, determined independently in the two arms of the chromosome. Reversals of direction

may occur at the kinetochore (interbrachial) or interstitially within an arm (intrabrachial). The present paper will be concerned only with interbrachial reversals and will deal with chromatid relational coiling proximal to the kinetochore, regardless of the direction of twisting in more distal regions.

The direction of relational coiling has been reported to be random in the two arms of a chromosome in *Trillium*,<sup>1, 2</sup> *Vicia*,<sup>1</sup> *Hyacinthus*<sup>3</sup> and *Allium*<sup>4</sup> and non-random in *Tradescantia*<sup>5</sup> and in the *B* chromosome of *Trillium*.<sup>2</sup> Koshy states that reversals of direction always occur at the kinetochore in *Allium*<sup>6</sup> and *Aloe*<sup>7</sup> and considers them essential to relational untwisting. Darlington<sup>8</sup> found that the direction of relational coiling "shows no consistency" in *Spironema*, *Crocus* and *Pushkinia*, is "sometimes inconsistent" in particular arms in *Nomocharis* and is "usually consistent in each arm of all chromosomes in *Fritillaria* and constant for the identifiable *M* chromosomes."

In *Trillium* the major coil of the first meiotic division persists throughout second division and appears as a plectonemic relic coil at prophase of the first division of the microspore.<sup>2, 9</sup> The relational coil is derived directly from the relic coil and, therefore, should show the same randomness in direction of coiling as does the major coil of meiosis.<sup>10, 11, 12, 13</sup> However, in an analysis of chromatid relational coiling in *Trillium* microspores Sparrow, Huskins and Wilson found that coiling was not random in the *B* chromosome, although it was random in the *C*, *D* and *E* chromosomes. (The kinetochore in the *A* chromosome is subterminal.)

This brief review of the literature on relational coiling indicates much disagreement. The present paper is an attempt to reconcile some of the apparent discrepancies by a detailed study of relational coiling in a plant for which both random and non-random coiling have been reported.

*Methods.*—Fresh anther smears of *Trillium grandiflorum* Salisb. were stained with iron aceto-carmin, pressed and sealed in the usual manner. Some of the slides were made permanent but a few "temporary" mounts have been kept in cold storage for nearly three years and are still usable. Camera lucida drawings of whole complements were made from nuclei in which the chromosomes were well separated. The number of relational twists present in an arm decreases as prophase advances and by metaphase none or only a few remain. Early stages are difficult to analyze and late stages have too many untwisted arms. Consequently the majority of the cells studied were late prophases in which most of the arms had two or more twists. Their mean total chromatid length per complement was 158.1  $\mu$  with a Standard Error of 4.3  $\mu$ . (See previous publications<sup>2, 9</sup> for illustrations and data on chromatid length changes.)

*Results.*—The direction of relational twists proximal to the attachment region of the *B*, *C*, *D* and *E* chromosomes are given in table 1 for a total

of 89 cells. Table 2 gives the number of chromosomes with reversals (R.L), the number without reversals (R.R and L.L) and in the third column the remainder. The latter group includes those chromosomes in which at least one arm had lost all its relational twists plus three cases in which it was not possible to determine the direction of coiling of the most proximal twists.

If the direction of coiling were independent in the two arms of a chromosome then the number of chromosomes with interbrachial reversals should be approximately equal to the number without them. For the *E* chromosome the numbers do not deviate significantly from the theoretically expected 1:1 ratio ( $\chi^2 = 1.026$ ), but the *B*, *C* and *D* chromosomes all show a statistically significant deviation (see table 2 for  $\chi^2$  and *P* values). There is an excess of chromosomes without reversals. Apparently, therefore, the direction or relational coiling is random across the kinetochore in the *E* chromosome but not in the *B*, *C* and *D* chromosomes, or at least so it appears when only chromosomes with both arms still relationally twisted are considered. When all chromosomes are considered a different interpretation is possible.

TABLE 1

Direction of Relational Twists Proximal to the Kinetochore in the *B*, *C*, *D* and *E* Chromosomes of 89 Microspores of *Trillium grandiflorum*

(R = right, L = left, U = untwisted, . = kinetochore)

CHROMOSOME	R.R	R.L	L.L	U.R	U.L	U.U	?
<i>B</i>	23	17	12	10	23	4	0
<i>C</i>	16	10	9	21	20	12	1
<i>D</i>	18	20	22	13	7	8	1
<i>E</i>	11	44	24	5	4	0	1
—	—	—	—	—	—	—	—
Total	68	91	67	49	54	24	3

Relational twists are not eliminated from all chromosomes simultaneously. The shorter arms having fewer relational twists untwist sooner, on the average, than the longer arms. Thus the *B*, *C* and *D*, respectively, had 37, 53 and 28 chromosomes with one or both arms untwisted while the *E* had only nine (table 2). Since all the chromosomes are from the same 89 cells any difference in number of relational twists (or lack of them) must be due to morphological or structural differences in the chromosomes themselves. The most obvious difference is in the position of the kinetochore and the relative lengths of the chromosome arms. The *E* chromosome, with median kinetochore and both arms long, untwists, on the average, later than the chromosomes with shorter arms and sub-median kinetochore. Consequently the *E* chromosome data most nearly represent the initial relational twisting of a sample of cells in which untwisting has not yet begun.



Since a 1:1 ratio of chromosomes with interbrachial reversals to those without is expected for all chromosomes, then the *B*, *C* and *D* chromosomes show a significantly lower number of chromosomes without reversals. On the basis of selective or non-random untwisting this means that more chromosomes with reversals than without have untwisted. The remaining chromosomes no longer can be considered as representative of an unselected sample.

The limits of a 1:1 ratio can be calculated from the formula  $\sigma = \sqrt{npq}$ . When  $n = 89$  and  $p$  and  $q = 0.5$  the value of  $\sigma$  is 4.72. The limits of the 1:1 ratio are then  $44.5 \pm 2\sigma$  or 35.1 and 53.9. Of the 89 *B* chromosomes 37 have one or both arms untwisted and hence cannot be classified as to reversals at the kinetochore. However, if 19 of them had previously had reversals then 36 of the total of 89 would have reversed direction leaving 53 without reversals. These numbers are just within the limits of the 1:1 ratio. In order to approximate more closely to a perfect 1:1 ratio more than half of the untwisted chromosomes must be considered to have had an interbrachial reversal. The same is true for the *C* and *D* chromosomes. It is therefore suggested that the apparent excess of chromosomes without interbrachial reversals is the result of earlier untwisting in at least one arm of chromosomes with interbrachial reversals. It thus appears that uncoiling does not occur entirely independently in the two arms of a heterobrachial chromosome, but is affected by the presence or absence of interbrachial reversals.

TABLE 2

Chromosomes with and without Reversals at the Kinetochore and the  $\chi^2$  and *P* Values Corresponding to the Observed Deviation from the Expected 1 : 1 Ratio

CHROMOSOME	NUMBER WITH REVERSALS	NUMBER WITHOUT REVERSALS	NUMBER NOT DETERMINED	$\chi^2$	<i>P</i>
<i>B</i>	17	35	37	6.230	<0.02
<i>C</i>	10	25	54	6.428	<0.02
<i>D</i>	20	40	29	6.666	<0.01
<i>E</i>	44	35	10	1.026	>0.3
Total	91	135	130	$\Sigma\chi^2 = 20.350$	<0.01

In chromosomes devoid of reversals untwisting can occur only at the distal ends. Intrabrachial reversals allow direct cancellation and with interbrachial reversals untwisting may occur also at the proximal ends. If the kinetochore is undivided direct cancellation cannot take place at the interbrachial reversal but essentially the same result will be obtained if the kinetochore rotates. This would be expected to happen only when the untwisting forces from the two arms are augmentive as they would be when the two arms are twisted in opposite directions.

Intrabrachial reversals were previously shown to play a part in the

elimination of relational coiling.<sup>2</sup> It now appears that both intra- and interbrachial reversals are instrumental in reducing the number of chromatid relational twists during prophase uncoiling. Further, if interbrachial reversals in heterobrachial chromosomes lead to non-random uncoiling it is apparent that non-randomness in direction of relational coiling does not necessarily mean that the spirals were non-random at the time of their inception. It seems probable, therefore, that some of the previously reported non-randomness in direction of relational coiling has also been the result of the alteration of the original 1:1 ratio by disproportionate untwisting in chromosomes with interbrachial reversals. The extent of the deviation from the expected ratio would presumably increase in accordance with the amount of untwisting. A sample of late prophases would thus be expected to have a greater deviation than a sample composed largely of earlier prophases. Considerable variation from sample to sample should, therefore, be expected.

*Summary.*—Four chromosomes from each of 89 microspores of *Trillium grandiflorum* Salisb. were analyzed to determine whether or not the direction of chromatid relational coiling was random across the kinetochore. The direction of coiling was random in the *E* chromosomes but apparently non-random in the *B*, *C* and *D* chromosomes. In the latter the excess of chromosomes without interbrachial reversals is not to be regarded as an initial non-randomness attributable to inherent or mechanical factors. The evidence is taken to indicate not that the direction of coiling was non-random at its inception but that the non-randomness arose as a result of disproportionate untwisting in chromosomes with interbrachial reversals. In brief, the apparent non-random coiling is in reality a non-random untwisting. After complete untwisting in some of the chromosomes arms the direction of coiling across the kinetochore in the remaining heterobrachial chromosomes then shows a significant deviation from the expected randomness.

\* National Research Fellow.

<sup>1</sup> Sax, K., *Genetics*, 21, 324-338 (1936).

<sup>2</sup> Sparrow, A. H., Huskins, C. L., and Wilson, G. B., *Can. Jour. Res., C*, 19, 323-350 (1941).

<sup>3</sup> Upcott, M., *Cytologia*, 8, 398-407 (1938).

<sup>4</sup> Husted, L., *Ibid.*, 8, 368-376 (1938).

<sup>5</sup> Husted, L., *Genetics*, 21, 537-553 (1936).

<sup>6</sup> Koshy, T. K., *Jour. Roy. Mic. Soc. (Ser. 3)*, 53, 299-318 (1933).

<sup>7</sup> Koshy, T. K., *Ann. Bot. (N. S.)*, 1, 43-58 (1937).

<sup>8</sup> Darlington, C. D., *Cytologia*, 7, 248-255 (1936).

<sup>9</sup> Sparrow, A. H., *Can. Jour. Res., C*, 20, 257-266 (1942).

<sup>10</sup> Matsuura, H., *Jour. Faculty Sci. Hokkaido Imp. Univ. (Ser. 5)*, 3, 233-250 (1935).

<sup>11</sup> Matsuura, H., *Cytologia*, 8, 178-194 (1937).

<sup>12</sup> Huskins, C. L., and Wilson, G. B., *Ann. Bot. (N. S.)*, 2, 281-291 (1938).

<sup>13</sup> Wilson, G. B., and Hutcheson, I., *Can. Jour. Res., C*, 19, 383-390 (1941).



## TUMOR FORMATION BY ATTENUATED CROWN-GALL BACTERIA IN THE PRESENCE OF GROWTH-PROMOTING SUBSTANCES

BY ARMIN C. BRAUN AND THOMAS LASKARIS

DEPARTMENT OF ANIMAL AND PLANT PATHOLOGY, THE ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH, PRINCETON, N. J.

Communicated September 28, 1942

It has recently been shown<sup>1, 2</sup> that crown-gall tumor cells, when freed of the original inciting agent, *Phytomonas tumefaciens* (Smith and Town.), Bergey, *et al.*, are capable of autonomous development. The fact that these tumor cells when grown *in vitro* retain indefinitely their peculiar cultural and cytological characteristics as well as their tumor-inducing capacity in the absence of any demonstrable stimulating agent indicates that the host cells when acted upon by *P. tumefaciens* become permanently altered. They behave like malignant animal cells in that they (1) show uncontrolled growth in the initial host, (2) can be transplanted successfully to hosts of the same species from which they were derived or to closely related hosts and there continue their uncontrolled growth and (3) retain indefinitely their ability to produce tumors and to grow autonomously *in vivo*. The nature of the fundamental change which occurs in the cells is at present not known. However, a better understanding of the factor or combination of factors involved in tumor formation should be of aid in the clarification of the physiological basis for the development of this neoplastic growth.

In recent years a number of investigators have attempted to demonstrate a possible relationship between growth-stimulating substances and tumor development. Certain of these workers<sup>3, 4, 5, 6</sup> have drawn a parallel between the reaction of plant tissues treated with  $\beta$ -indole acetic acid, a substance produced by the crown-gall organism from tryptophane, and the reaction of similar plant material inoculated with a living culture of *Phytomonas tumefaciens*. Although Riker and his associates<sup>7</sup> have confirmed the experimental results of the above workers, they deny the importance of  $\beta$ -indole acetic acid or similar growth substances in tumor formation and conclude that, as far as they are aware, *P. tumefaciens* is pathogenic independently of these materials. Levine<sup>8</sup> agrees with Riker and states that no evidence has thus far been adduced to show that the mechanism of tumor formation is due to the presence of  $\beta$ -indole acetic acid.

The earlier work of Locke, Riker and Duggar<sup>9</sup> indicated, nevertheless, that there was a high level of growth-promoting substance present in plants inoculated with a virulent culture of the crown-gall organism. This was evidenced by marked growth responses such as epinasty of the leaf petioles, inhibition of the axillary buds, etc., that resulted, in addition

to tumor formation, in certain host plants. Plants inoculated with an attenuated culture show these growth responses to a much lesser degree or not at all. This would suggest that growth substances may be a limiting factor in tumor formation by the attenuated culture. Locke, *et al.*, were not able, however, to bring about an appreciable increase in the size of the tumors formed by supplementing the attenuated culture with  $\beta$ -indole acetic acid.

It is the purpose of the present paper to demonstrate that an attenuated culture of the crown-gall organism, when supplemented with certain growth-promoting substances, is capable of inducing the formation of large tumors in tomato plants. Evidence is presented, furthermore, that tissue fragments from such experimentally induced tumors retain undiminished their tumor-inducing capacity upon transplantation.

*Methods.*—The attenuated isolate of *Phytoplasma tumefaciens* used in these experiments was first described by Hendrickson, *et al.*,<sup>10</sup> and designated by them as the A66 strain. This isolate is capable of inducing cellular proliferation at the point of inoculation somewhat in excess of the normal wound response of tomato plants, but it is unable to cause the development of crown-gall tumors under ordinary conditions (Fig. 1 (b)). Available for comparative purposes was the A6 strain from which the attenuated A66 culture was derived. The A6 strain is highly virulent, as shown by its ability to incite the formation of massive galls at points of inoculation (Fig. 1 (a)).

The host plant used throughout the tests was the tomato (*Lycopersicon esculentum* Mill. var. Bonny Best). Young seedlings were transplanted to 4-inch pots and kept on a greenhouse bench throughout the course of an experiment. When the plants reached a height of approximately 15 inches, they were decapitated and inoculated by means of the needle-puncture method with the attenuated A66 culture. Two inoculations were made in each plant at distances of approximately  $\frac{1}{4}$  and  $\frac{3}{4}$  inch below the decapitated stem. Four days after inoculation, chemically pure growth substances in concentrations of 0.5, 1.0, 1.5 and 2.0% in lanolin were applied separately to the decapitated stumps of the inoculated plants. At intervals of 7 days, the growth substances were renewed until 3 applications had been made. The growth substances used included  $\alpha$ -naphthalene acetic acid,  $\beta$ -indole acetic acid and  $\gamma$ -indole butyric acid. Controls for the above experiments were set up as follows: One series of decapitated plants was punctured with a sterile needle instead of being inoculated with the attenuated culture but received the corresponding growth-substance applications; in a second series, pure lanolin was applied to plants inoculated with the attenuated culture.

The following technique, successfully employed by us in the tissue transplantation experiments to be described later, was adopted after a number

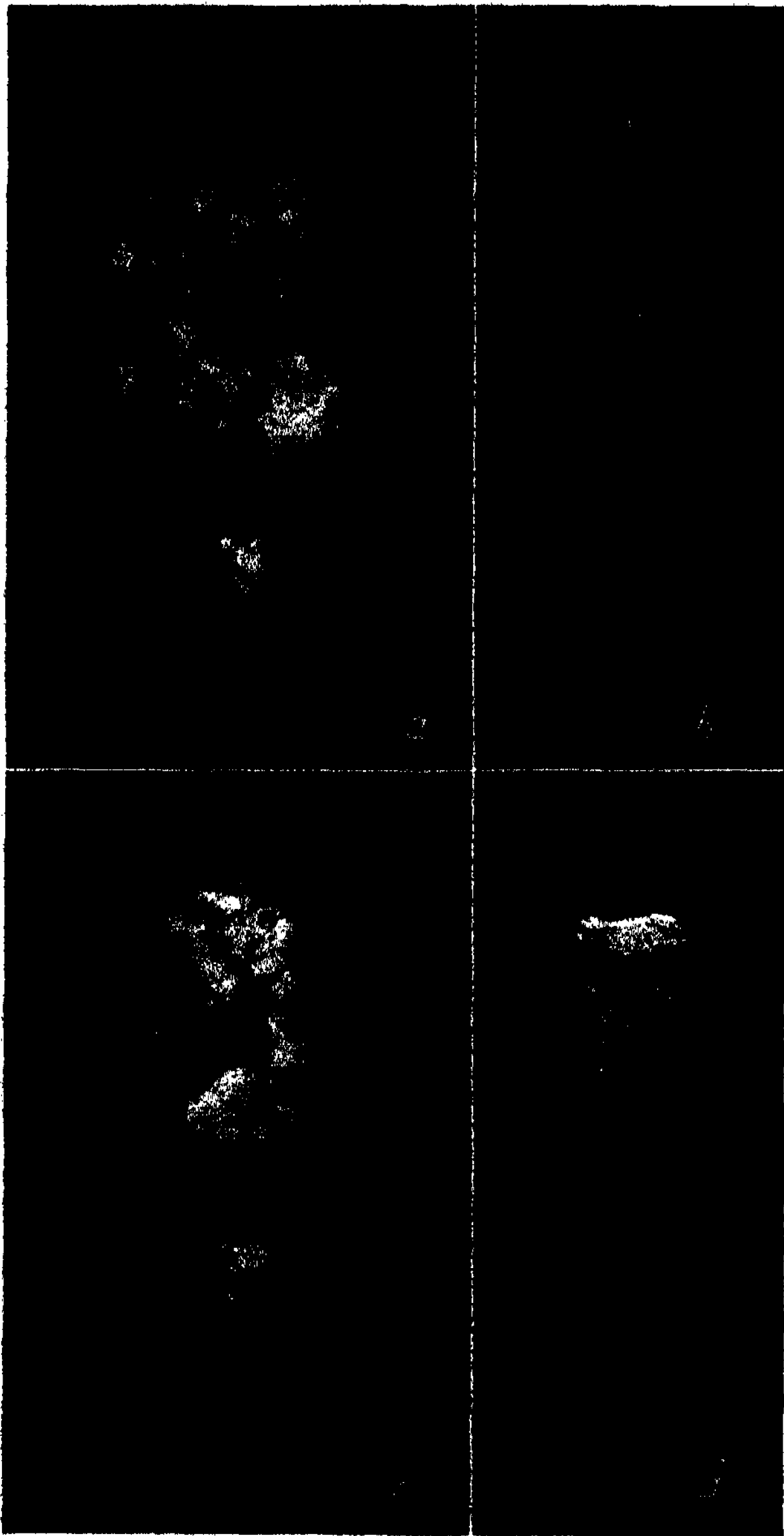


FIGURE 1

Decapitated tomato plants inoculated with (a) the virulent A6 culture, (b) the attenuated A66 culture, (c) the attenuated A66 culture and treated with synthetic growth-promoting substance. (d) Uninoculated, punctured with a sterile needle and treated with synthetic growth-promoting substance. Note similarity of tumors formed in (a) and (c). (Photographs by J. A. Carlile.)

of preliminary experiments established its superiority over the common grafting procedures. With the aid of a sharp scalpel, two parallel incisions were made along the axis of the stem, usually well above the middle of the plant. The incisions were approximately  $\frac{1}{4}$  inch apart and 1 inch in length. By careful insertion of the scalpel blade between the two incisions, the epidermal strip, together with the subjacent cortical and phloem elements, was pried loose from the main body of the stem and the desired tumor tissue fragment was then placed in the space between the flap and the stem. The implanted tissue was wrapped securely with "Sterilastic" tape which was removed 12 to 14 days later. The successful development of the implant appeared to be dependent upon early and satisfactory vascularization between the host and the tissue fragment. Tissues inserted inside or outside the confines of the vascular ring failed to develop in the majority of cases.

*Experimental Results.*—The three growth-promoting substances used,  $\alpha$ -naphthalene acetic acid,  $\gamma$ -indole butyric acid and  $\beta$ -indole acetic acid were all found to have a pronounced stimulating effect causing the development of large tumors in plants that had been previously inoculated with the attenuated A66 culture. However, consistent differences in the degree of stimulation, rate of tumor development and morphology of the resulting tumor mass were exhibited by the growth-promoting substances. The stimulating action of  $\gamma$ -indole butyric acid was usually observed somewhat earlier than that of  $\alpha$ -naphthalene acetic acid. The latter, however, gave rise to tumors that most closely resembled those initiated and stimulated by the virulent A6 culture. These tumors were white in color and irregular in contour, while those produced with the aid of  $\gamma$ -indole butyric acid were regular in contour and had a decidedly greenish tinge. In most instances the tumors stimulated by these two substances grew rapidly and became very large. Sizable tumors frequently developed after 3 applications of 0.5% of these growth substances in lanolin or following a single application of 1.0%.  $\beta$ -Indole acetic acid was not as effective in our hands as were the other two materials. Higher concentrations, usually 3 applications of 2.0% in lanolin, were required to bring about a tumor stimulation comparable to that secured with lower concentrations of the other two substances. The  $\beta$ -indole acetic acid-stimulated tumors, however, frequently grew to be very large and resembled closely those stimulated by  $\gamma$ -indole butyric acid. Figures 1 (c), 2 (a) and 2 (b) show, respectively, the stimulatory effects of 3 applications of 1.0% of  $\alpha$ -naphthalene acetic acid, 2.0% of  $\beta$ -indole acetic acid and 1.0% of  $\gamma$ -indole butyric acid in lanolin. The control, an attenuated culture-inoculated plant treated with pure lanolin, is shown in figure 1 (d).

The stimulatory effects of these growth substances on tumor development by the attenuated culture were not always evident even though

conditions were favorable. Thus, under completely comparable conditions, tumors on some plants would be stimulated to further development while not on others. The explanation of this variability may involve differences in rate of diffusion of growth-promoting substances in different plants.

The action of  $\alpha$ -naphthalene acetic acid,  $\beta$ -indole acetic acid and  $\gamma$ -indole butyric acid on tomato plants has been studied by many investigators in recent years. In our work as well as in the work of others it has been found that these substances in the concentrations commonly used are

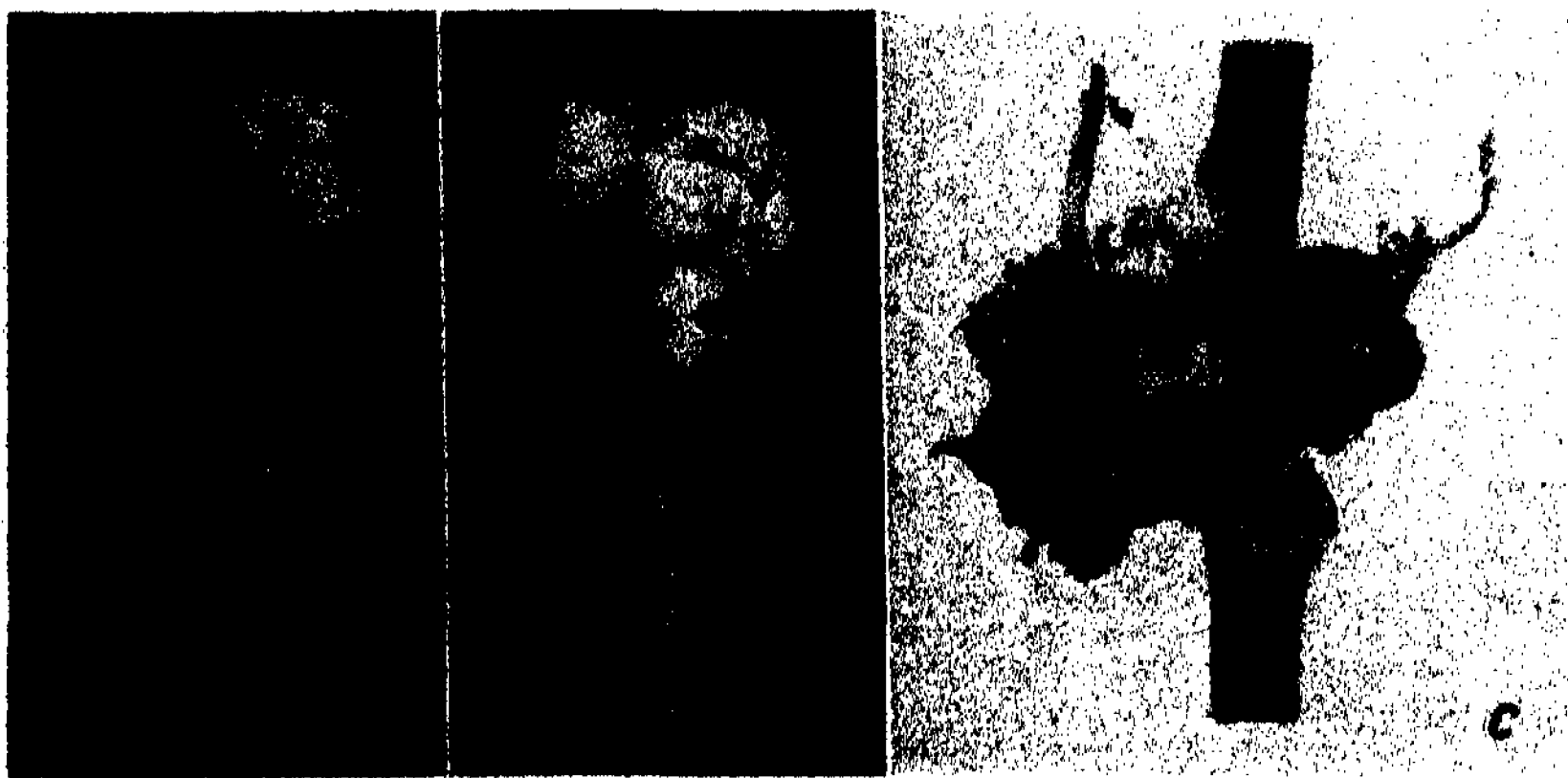


FIGURE 2

Decapitated tomato plants inoculated with the attenuated A66 culture and stimulated with (a)  $\beta$ -indole acetic acid, (b)  $\gamma$ -indole butyric acid and (c) the plant growth hormones. (Photographs by J. A. Carlile.)

essentially root stimulants, although some hypertrophied and hyperplastic tissues are formed. The latter present a histological picture not unlike that of crown-gall tumor tissue. The hypertrophied and hyperplastic cells are, however, endowed with very limited powers of proliferation and continue to develop only as long as they are kept stimulated by the growth substances. We have never observed overgrowths to be produced on tomato plants even by repeated applications of these substances that are comparable either in size or rate of development to those produced by the crown-gall organism. Figure 1 (a) shows the typical reaction of a decapitated tomato stem which had received 3 applications of 1.0% of  $\gamma$ -indole butyric acid in lanolin. A limited amount of cellular proliferation has occurred at the cut surface where the growth substance was applied; below the cut surface there are found numerous root primordia which under humid conditions would develop into true roots. The action of

$\alpha$ -naphthalene acetic acid and  $\beta$ -indole acetic acid on decapitated tomato stems is very similar.

The discovery that synthetic growth substances were able to stimulate the development of tumors by the attenuated culture strengthened our previous belief regarding the probable rôle of the host growth hormones in the development of these neoplastic growths. It had been observed that the removal of axillary shoots from decapitated tomato plants inoculated with the attenuated A66 culture frequently resulted in forcing the development of small adventitious buds on or in the proximity of the proliferating tissue. The further expansion and development of these buds were often accompanied by an expansion in the mass of the subjacent tumor tissue which increased at a tremendous rate and finally reached extremely large dimensions (Fig. 2 (c)). Because expanding buds are known to be a rich source of growth substances, it appears likely that the plant hormones secreted by the buds served as the stimulating agent in much the same manner as did the synthetic growth-promoting materials described above. No such stimulation was observed in those cases where the adventitious buds failed to develop. There is little question but that these "leafy galls" are composed largely of tumor tissue. Our evidence for this claim will be advanced in a later section.

The possibility that the application of the growth substances had significantly altered the degree of virulence of the attenuated culture was also investigated. Bacterial isolations were made from many of the artificially stimulated tumors by means of the usual poured-plate procedures. Large numbers of the bacterial colonies that developed were reinoculated into healthy tomato plants. In addition, tumor tissue fragments were thoroughly ground up with a small amount of nutrient dextrose broth in a mortar and the resulting suspension inoculated directly into healthy tomato plants. The results from many such tests demonstrated conclusively that the attenuated culture had not increased in virulence. It still produced only slight proliferation comparable to that produced by the stock A66 culture. The conclusion seems justified, therefore, that the primary effect of the growth substances was to stimulate the host cells rather than to bring about a change in the virulence of the bacteria themselves. Additional evidence in support of this conclusion is presented below.

*Transplantability of Tumor Tissue.*—Perhaps the most distinctive characteristic of malignant cells is their relative independence of the conditions which govern the growth of normal tissues. All normal cells are subject to a rigid control mechanism when growing *in vivo*, and it is only transformed cells such as are found in neoplasms that react differently. These abnormal cells fail to respond to the morphogenetic restraints which direct the growth of cells in a healthy organism. As was pointed out earlier,

crown-gall tumor cells, like most malignant animal cells, appear to be of a type which is capable of unlimited transplantability and tumor-inducing capacity under favorable conditions. We have used, therefore, these criteria in determining whether the attenuated A66 culture of *Phytomonas tumefaciens* is capable, as is the virulent culture, of permanently altering the host cells.

Fragments from tumors initiated by the attenuated culture and stimulated with  $\alpha$ -naphthalene acetic acid,  $\gamma$ -indole butyric acid and the natural plant growth hormones were implanted under the epidermis of tomato plants. It was soon observed that in those cases where a vascular connection with the host was clearly established the tissue fragments developed into typical crown-gall tumors in relatively short periods of time. This was especially true of the  $\alpha$ -naphthalene acetic acid and  $\gamma$ -indole butyric acid-stimulated tissue which in many instances developed into large tumors 3 to 4 cm. in diameter within a period of 4 to 5 weeks. On the other hand, the tissue fragments from the plant growth hormone-stimulated "leafy" galls developed more slowly and usually did not exceed 1.5 to 2 cm. diameter in 5 weeks of development.

As indicated above, the gross morphological characteristics of the tumors initiated by the attenuated culture and stimulated by various growth substances differed considerably. However, upon transplantation of the various tumors, all of the resulting overgrowths appeared similar and could not be distinguished from each other except possibly by their respective rates of development. The tumors were for the most part a very light brown in color and of a fairly regular contour.

We have now successfully carried some of these tumors through 5 successive passages in tomato plants over a period of more than 6 months. These tumors have continued to develop at an undiminished rate throughout this period. Figure 3 (a) shows a typical third-generation tumor transplant after 4 weeks of development.

Additional bacterial isolations were made from these transplanted tumors to determine whether the degree of virulence of the attenuated culture had been in any way altered during the course of the experiments. It was found after testing several hundred isolates on tomato plants that no increase in virulence over that of the stock A66 culture had occurred. It seemed logical to assume, therefore, that the mere presence of these attenuated organisms in the tissues could not account for the rapid expansion of the transplanted fragments and that it must be the host cells themselves that had been altered to such a degree that they were now developing autonomously. This assumption was definitely established by our subsequent discovery that certain tissue fragments of the  $\gamma$ -indole butyric acid-stimulated tumors were entirely free of *Phytomonas tumefaciens*. Whether the bacteria had died out in these fragments or whether



the rapidly developing tumor tissue had grown away from the bacteria is not known. The bacteria-free tissues have been carried through 3 transplant generations involving a period of more than 3 months. All attempts to demonstrate the presence of the crown-gall bacteria in these

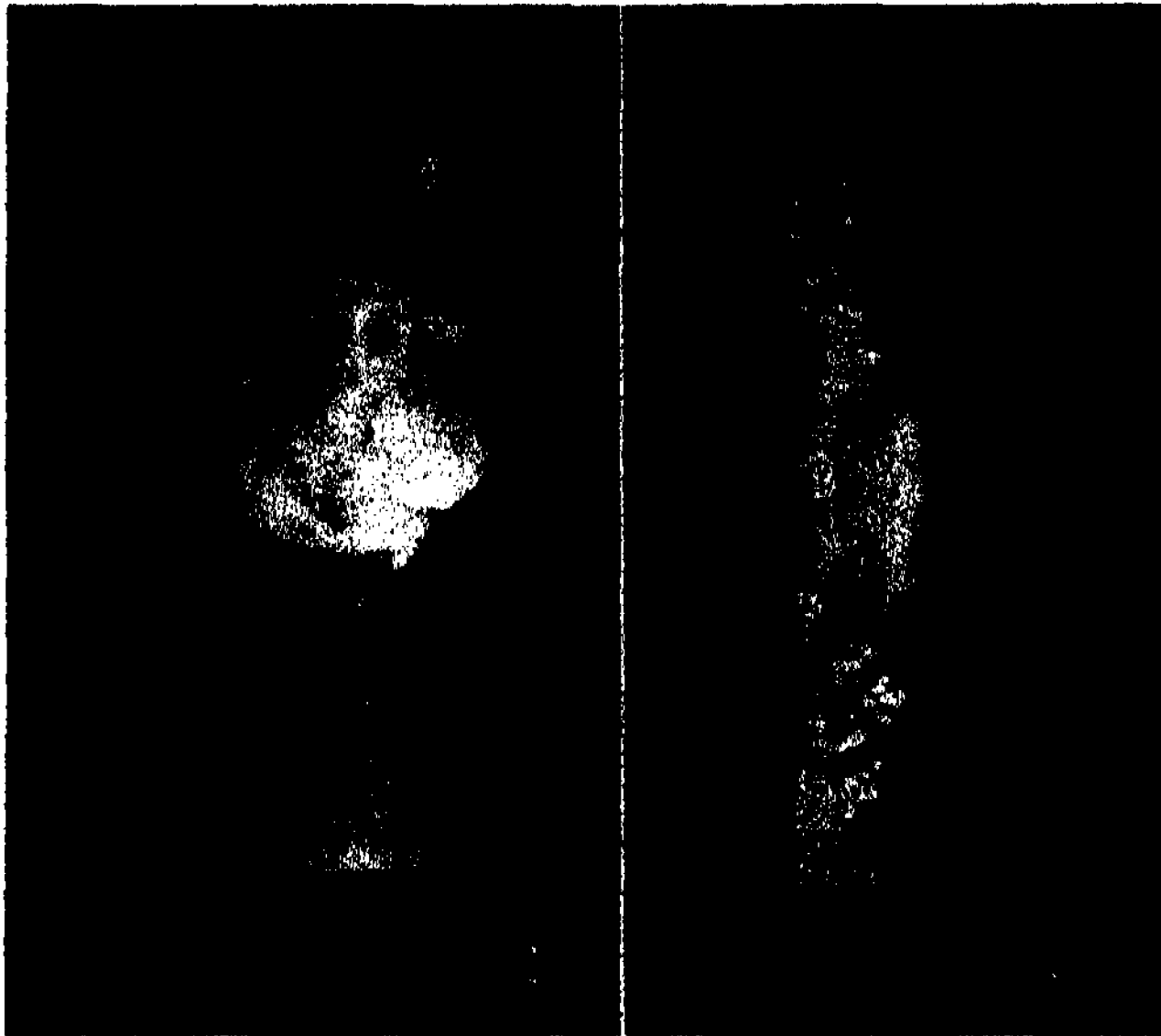


FIGURE 3

(a) Typical third-generation implant (originally derived from an experimentally induced tumor, Fig. 1 (c)) after 4 weeks of development. (b) Typical first-generation implants of tissue stimulated with growth-promoting substance (see Fig. 1 (d)) after 6 weeks of development. Two general types of response are illustrated. In the upper implant the tissue remained smooth and made but slight growth. The nodular root-like protuberances and limited growth of the lower implant characterized the second type of response. (Photographs by J. A. Carlile.)

tissues by the usual isolation procedures have failed. The bacteria-free tumor tissue fragments grew rapidly and developed into large neoplastic growths indistinguishable morphologically from the bacteria-containing tumors. On the basis of these results, it can be stated, therefore, that the attenuated bacteria are capable of altering the host cells to tumor cells which after stimulation are apparently capable of autonomous development.

The controls for the experiments described in the preceding pages were uninoculated decapitated tomato plants treated with 1.0, 1.5 and 2.0%



of  $\beta$ -indole acetic acid and  $\alpha$ -naphthalene acetic acid in lanolin, as explained earlier. These materials were found to be essentially root-forming substances, although they did cause a limited amount of cellular proliferation. Fragments of these growth substance-stimulated tissues were grafted into tomato plants in the same manner as were the tumor tissue fragments. Similar grafting experiments were carried out on three separate occasions with the object of determining whether the growth substances used here were capable of permanently altering the host cells. In no instance did we find evidence suggesting that such a change had occurred. The tissue implants seldom increased to more than twice their original size and never developed into tumors. Two general types of responses were observed (Fig. 3 (b)). In one type of response the tissue implant gave rise to numerous nodular root-like protuberances which seldom developed into true roots, although occasionally roots 1 to 2 cm. in length were observed. The second type could be likened to a slow healing response, the inserted tissue remaining smooth and making but slight growth. It is apparent that the host cells, when acted upon by these growth substances alone, are incapable even after successful transplantation of inducing the formation of tumors. It is our belief that the growth substances used in this study served merely to stimulate cells previously altered by the attenuated culture.

*Discussion and Conclusions.*—It has been demonstrated that an attenuated culture of *Phytomonas tumefaciens* is capable of forming large tumors in tomato plants when supplemented with the plant growth hormones or synthetic growth-promoting substances. This growth substance reaction does not appear to be specific inasmuch as any one of several different growth substances, including that produced by the host plant, may serve as the stimulating agent. Furthermore, the gross morphological appearance of the tumors formed when an attenuated culture is supplemented with a growth substance varies with the substance used. Of such tumors, those stimulated by  $\alpha$ -naphthalene acetic acid approximate most closely those induced by the virulent culture on tomato in that they are white in color and irregular in contour.

It has further been demonstrated by means of grafting experiments that the cells of these experimentally induced tumors can be transplanted in series and subsequently develop into large tumors which may reach a size of 3 to 4 cm. in a period of 5 weeks. Since certain of these tumors are apparently bacteria-free and since it has been shown in the case of those tumors which are not bacteria-free that the degree of virulence of the attenuated culture itself has not been altered by the application of growth substances, it is concluded that the attenuated culture like the virulent culture is capable of altering the host cells to tumor cells. However, the attenuated culture alone is unable to stimulate the altered cells to any

appreciable extent. When artificially stimulated with growth substances, these altered cells are apparently capable, as are those stimulated by the virulent culture, of uncontrolled growth *in vivo*. Thus, there appear to be at least two distinct phases involved in tumor formation. In the first phase, the normal host cells are changed to tumor cells which without stimulation will not develop into a neoplastic growth. The second phase consists in the stimulation of the changed cells to continued multiplication by a growth substance, resulting ultimately in the formation of a tumor.

The question as to whether a single substance or two or more substances are involved in the alteration and stimulation of the tumor cells remains as yet unanswered. If only a single substance is involved, however, it must differ from the growth substances used here because the cells of the tomato plant stimulated by these substances alone do not, upon transplantation, induce the formation of tumors. If two or more substances are involved, then one of the common growth-promoting substances might conceivably act as the stimulating agent. The difference between the virulent and the attenuated cultures may lie in the relative amounts of growth substances produced by each, either as a product of bacterial metabolism or by the host under the influence of these organisms. The fact that pronounced growth substance responses are characteristically shown by tomato plants inoculated with the virulent culture and are absent in similar plants inoculated with the attenuated culture lends credence to this belief.

Once the altered host cells are sufficiently stimulated, they are apparently capable of indefinite multiplication and tumor formation under favorable conditions without the additional application of growth-promoting substances. The altered cells then continue to multiply autonomously and retain, perhaps indefinitely, their tumor-inducing capacity.

<sup>1</sup> Braun, A. C., and White, P. R., *Phytopath.* (in press).

<sup>2</sup> White, P. R., and Braun, A. C., *Cancer Res.*, **2**, 597-617 (1942).

<sup>3</sup> Berthelot, A., and Amoureux, G., *C. R. Acad. Sci. Paris*, **206**, 537-540 (1938).

<sup>4</sup> Brown, N. A., and Gardner, F. E., *Phytopath.*, **26**, 708-713 (1936).

<sup>5</sup> Kraus, E. J., Brown, N. A., and Hamner, K. C., *Bot. Gaz.*, **98**, 370-420 (1936).

<sup>6</sup> Link, G. K. K., Wilcox, H. W., and Link, A. DeS., *Ibid.*, **98**, 816-867 (1937).

<sup>7</sup> Riker, A. J., Henry, B., and Duggar, B. M., *Jour. Agr. Res.*, **63**, 395-405 (1941).

<sup>8</sup> Levine, M., *Bull. Torrey Bot. Club*, **67**, 199-226 (1940).

<sup>9</sup> Locke, S. B., Riker, A. J., and Duggar, B. M., *Jour. Agr. Res.*, **57**, 21-39 (1938).

<sup>10</sup> Hendrickson, A. A., Baldwin, I. L., and Riker, A. J., *Jour. Bact.*, **28**, 597-618 (1934).

## THE RÔLE OF INTRACELLULAR CATIONS ON LIVER GLYCOGEN FORMATION *IN VITRO*

BY A. BAIRD HASTINGS AND JOHN M. BUCHANAN

DEPARTMENT OF BIOLOGICAL CHEMISTRY, HARVARD MEDICAL SCHOOL

Communicated October 14, 1942

The importance of specific inorganic ion concentrations in the extracellular fluids of the body for the maintenance of the integrity and normal function of cells has long been recognized. The development of "balanced" salt solutions, such as Ringer's solution, as a substitute for natural extracellular fluid is an illustration of this. In recent years, attention has turned, however, to the importance of the intracellular ion concentrations for the maintenance of normal intracellular enzyme activity.<sup>1,2</sup> This becomes of special urgency in *in vitro* studies in which metabolic functions of tissues are under investigation. For example, the intracellular fluid of mammalian liver is high in potassium and magnesium and low in (or perhaps free from) sodium and calcium; whereas the extracellular fluid of the tissue is similar to blood plasma, high in sodium and low in potassium, magnesium and calcium. When liver tissue is incubated *in vitro* at body temperature in an artificial extracellular medium such as Ringer's solution, there is a progressive change in permeability of the cells with a resultant exchange of intracellular and extracellular ions until equilibrium is established. Such an exchange of ions results in an intracellular ionic environment which is quite different from that which normally obtains in the body.

However, by placing tissues in a fluid similar to the intracellular fluid with respect to its cations, it should be possible to maintain the normal intracellular cation environment in spite of exchange of ions across the cell boundary.

The present communication reports the effects the differences in ionic environment produced upon one metabolic function of rat liver—namely, the production of glycogen from glucose *in vitro*. Our experiments were stimulated by the work of Ostern, Herbert and Holmes<sup>3</sup> who reported in 1939 the *in vitro* production of glycogen by rabbit liver when incubated with glucose in Ringer's solution enriched with calcium. These experiments with rabbit liver were repeated and confirmed in our laboratory, but rat liver, under the same conditions, failed to produce significant amounts of glycogen. We, therefore, turned to a more critical examination of the effect of intracellular versus extracellular cations on glycogen formation.

The specific problem undertaken was the comparison of the glycogen content of liver slices before and after they had been incubated for one hour at 38°, in isotonic glucose-containing solutions: (a) similar to intracel-

lular fluids with respect to cations—i.e., high in potassium and magnesium, and (b) similar to extracellular fluids—i.e., high in sodium with small amounts of potassium, calcium and magnesium. Comparable experiments were also performed in the absence of glucose. These provided information on the rôle of cations on net glycogen formation and breakdown. In view of the complexity of the intracellular organic phosphate anions, it is impractical at present to provide a normal intracellular anionic environment in the extracellular medium.

It may be stated at once that incubation of the liver in the high potassium, high magnesium medium favored glycogen synthesis, whereas incubation in the high sodium medium favored glycogen breakdown.

#### *Experimental.*

SOLUTION	Na, mM/L	K, mM/L	Ca, mM/L	Mg, mM/L	Cl, mM/L	HCO <sub>3</sub> , mM/L	GLUCOSE %
I	0	130	0	20	130	40	1.0
II	152	5	1	1	121	40	1.0
III	0	130	0	20	130	40	0
IV	152	5	1	1	121	40	0

The above solutions were equilibrated with 5% CO<sub>2</sub>:95% O<sub>2</sub>. When 2 cc. of such solutions were incubated in 25 cc. vessels with 300–500 mg. of rat liver slices for one hour, the resultant pH was found to be between 7.35 and 7.45. The glycogen content of liver plus the solution of each vessel was determined by the method of Good, Kramer and Somogyi<sup>4</sup> as modified by Sjörgen.<sup>5</sup> The total phosphorus content of each vessel was also determined by the method of Fiske and Subbarow.<sup>6</sup> The initial glycogen and phosphorus contents of the liver were determined before incubation in order to provide a basis for comparison. It was found to be more convenient to estimate the amount of liver in each vessel from the phosphorus determinations than to depend upon the weights of the slices of liver which was introduced. Results were calculated in terms of grams of glycogen per 100 grams of wet liver. The rats were starved for 18 hours prior to experimentation.

*Results.*—In 37 out of 42 experiments in which rat liver, whose initial glycogen was less than 0.3 per cent, was incubated with Solution I (i.e., high K and Mg), there was an absolute increase (average 0.05–0.15%) in the amount of glycogen present at the end of one hour. In experiments in which the initial glycogen was greater than 0.3 per cent, a decrease in glycogen was usually observed. It was concluded from these experiments that, contrary to the previous experience in which little or no increase in glycogen occurred when Ringer's solution was used, rat liver would quite consistently produce glycogen *in vitro* when incubated in a potassium-mag-

nesium medium providing the initial glycogen concentration did not exceed 0.3 per cent.

The results obtained in seven experiments in which the effects of using Solutions I, II, III, IV were compared are shown graphically in figures 1

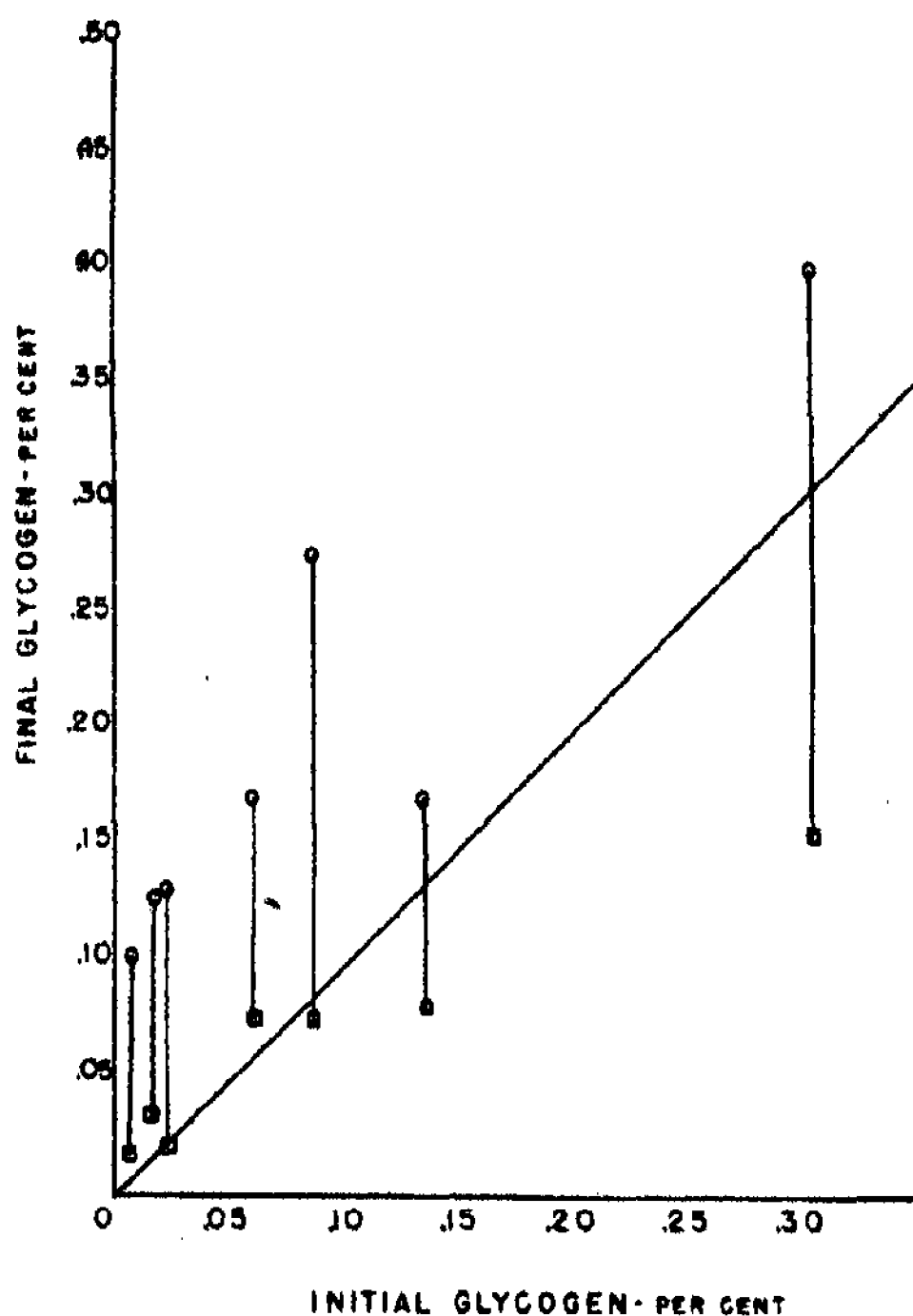


FIGURE 1

Final liver glycogen concentration is plotted against initial glycogen concentration. Circles designate experiments with Solution I (high potassium and magnesium, containing glucose); squares designate experiments with Solution II (high sodium containing glucose). Points lying below the diagonal line indicate net glycogen breakdown, points above the line indicate net glycogen formation.

and 2. In figure 1, it will be seen that, in the presence of glucose, the final glycogen concentration was absolutely greater than the initial concentration when the high potassium-magnesium medium (I) was used; whereas, when the high sodium medium (II) was used, the final glycogen concentration was less than or essentially the same as the initial concentration. Figure 2 shows that, in the absence of glucose, there was no loss of glycogen

in the high potassium-magnesium medium (III), whereas, the glycogen practically disappeared in the high sodium medium (IV). These results may be interpreted as indicating that high sodium and low potassium, cal-

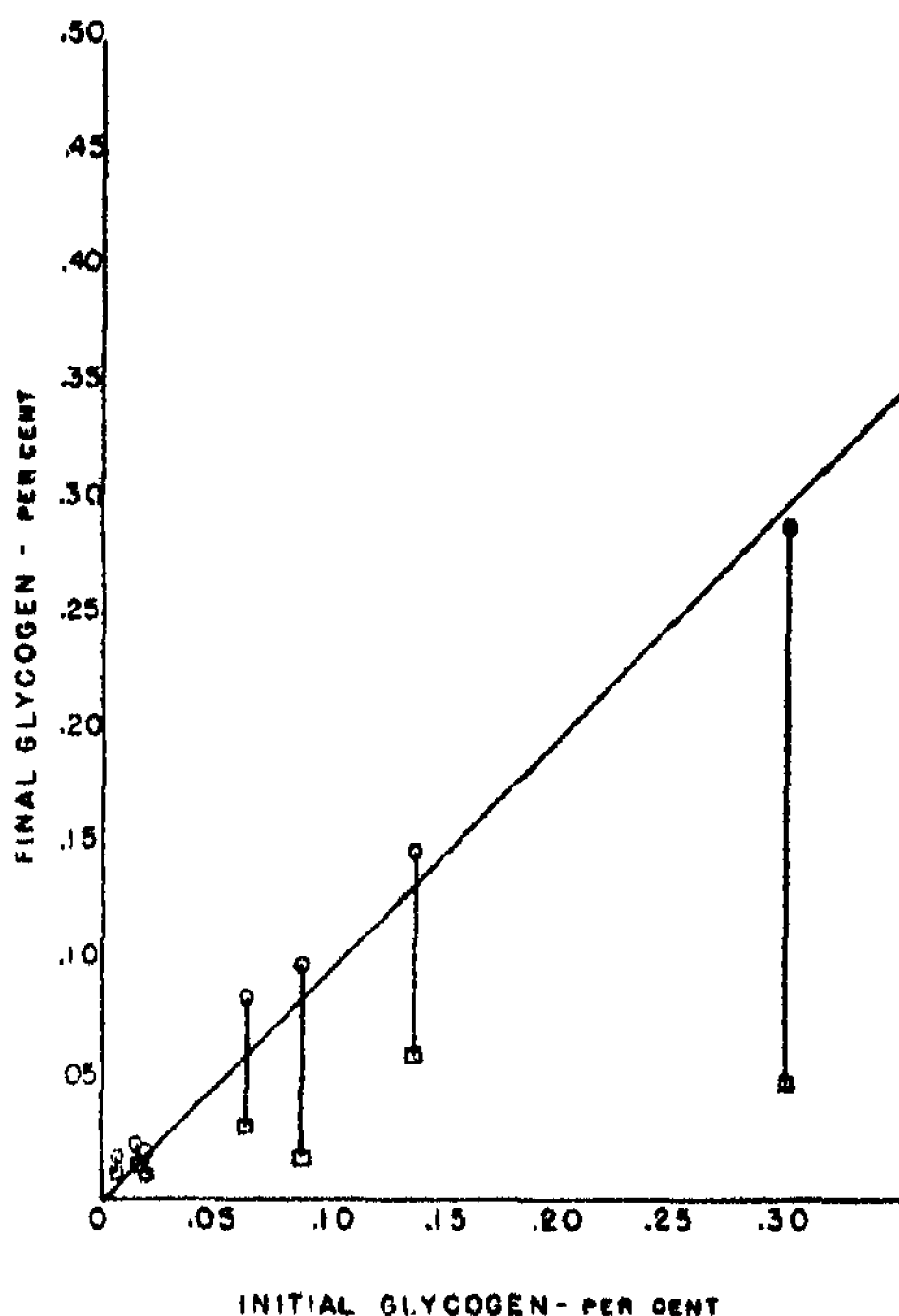


FIGURE 2

Final liver glycogen concentration is plotted against initial glycogen concentration. Circles designate experiments with Solution III (high potassium and magnesium, without glucose); squares designate experiments with Solution IV (high sodium, without glucose). Points lying below the diagonal line indicate net glycogen breakdown, points above the line indicate net glycogen formation.

cium and magnesium favored glycogen breakdown in excess of glycogen synthesis, whereas, the high potassium and high magnesium solution favored glycogen synthesis in excess of glycogen breakdown.

The above observations are to be regarded as only preliminary to a more detailed and complete analysis of the role of ions in influencing intracellular reactions. For example, it has already been observed in this laboratory by one of us (J. M. B.) that calcium is more effective than magne-

sium in promoting glycogen synthesis in rat liver *in vitro*. It is not the intention of the present communication to consider the question of what solution would provide for optimum rate of glycogen formation, but rather to call attention to the importance of considering the maintenance of the normal intracellular ionic environment in the study of intracellular enzymic reactions.

*Summary.*—1. The presence of concentrations of potassium and magnesium comparable to those occurring in intracellular fluid has been found to favor glycogen formation by rat liver *in vitro*, whereas incubation of liver in media comparable to extracellular fluid favors glycogen breakdown.

2. The importance of maintaining a normal intracellular ionic environment in the study of intracellular reactions is pointed out.

<sup>1</sup> Fenn, W. O., *Physiol. Rev.*, **20**, 377 (1940).

<sup>2</sup> Hastings, A. B., *Harvey Lectures*, **36**, 91 (1940–1941).

<sup>3</sup> Ostern, P., Herbert, D., and Holmes, E., *Biochem. Jour.*, **33**, 1858 (1939).

<sup>4</sup> Good, C. A., Kramer, H., and Somogyi, M., *Jour. Biol. Chem.*, **100**, 485 (1933).

<sup>5</sup> Sjögren, B., *Arch. ges. Physiol.*, **250**, 427 (1938).

<sup>6</sup> Fiske, C. H., and Subbarow, Y., *Jour. Biol. Chem.*, **66**, 375 (1925).

## ON THE MATERIAL EJECTED FROM NOVAE

BY CECILIA PAYNE-GAPOSCHKIN AND SERGEI GAPOSCHKIN

HARVARD COLLEGE OBSERVATORY

Communicated October 16, 1942

The extremely complex spectroscopic problems connected with the outbursts of novae present many angles of approach. The one that has been most thoroughly exploited is concerned with the identification of the several absorption and emission spectra, moving with different radial velocities through the ejected envelope. The problem cannot, however, be regarded as solved until not only the motion and distribution, but also the quantity and composition of the ejected matter has been analyzed. The present paper is concerned with an approach to the second of these questions.

In studying the changes in the spectra of novae during and after the outbursts it has been usual to estimate and describe the intensities of the dark and bright lines as seen against the continuous background of the spectrum. Such estimates, however, are deceptive, since the intensity of the background is itself changing rapidly (and not necessarily uniformly), so that the true changes, especially in the intensities of bright lines, may be falsified.

Such lines are often described as increasing in intensity with time, but when true intensities are measured, it is found that increases of intensity for bright lines in the spectra of novae are very uncommon.

1. *The Material.*—The intensities in the bright lines of novae have been under investigation for a number of years, principally on the basis of the Harvard collection of objective prism spectra.<sup>1</sup> The problems concerned with the use of photometrically unstandardized spectra, and the methods used for the determination of intensities, are sufficiently described in the papers cited. The intensities of the more prominent bright lines in the spectra of five bright novae (Nova Persei 1901, Nova Aquilae 1918, Nova Cygni 1920, Nova Pictoris 1925 and Nova Herculis 1934) are given for several months after the first outbursts. The object of the present note is to summarize the results of these investigations, and to draw some elementary conclusions about the quantity and condition of the material ejected.

2. *Summary of Results.*—The lines of hydrogen, helium (neutral and ionized), ionized carbon, twice ionized nitrogen and ionized iron are the principal permitted radiations shown by novae, and measures of all of these have been made for at least some of the stars. Furthermore, the "forbidden lines" of once ionized nitrogen, of neutral and twice ionized oxygen, of once ionized sulphur and of twice ionized neon, have been measured. It must be remembered that the bright lines of novae, like the absorption lines, are of complex structure, and that the measures refer to the whole line (undoubtedly formed at a variety of spots in an inhomogeneous, expanding envelope); quantitative results drawn from them are therefore only approximate, and refer essentially to conditions averaged over the volume from which the given lines emanate.

The measured intensities of the bright lines (which are fully tabulated, day by day, in the papers cited) are determined in ergs per square centimeter per second at the surface of the earth. The greatest intensities recorded are about  $10^{-6}$  ergs per second per square centimeter; the smallest intensities attainable with our material are about  $10^{-10}$  ergs per second per square centimeter. These quantities have little significance in comparing the data for different novae, which are at very different distances; a physical discussion requires a knowledge of the energy emitted per second *at the nova*. The necessary factors are derived in table 1.

TABLE 1  
REDUCTION TO ABSOLUTE INTENSITY

NOVA	ABSOLUTE VISUAL MAGNITUDE	MODULUS	LOGARITHM OF REDUCTION FACTOR
Per 1901	— 8.7	8.7	43.56
Aql 1918	—10.0	8.6	43.52
Cyg 1920	— 8.7	10.7	44.34
Pic 1925	— 7.7	8.9	43.64
Her 1934	— 7.1	8.5	43.48



For comparison, the absolute intensities radiated in the hydrogen line  $H\delta$  by the five different novae at a number of different absolute magnitudes are shown in table 2. The intensity passes through a flat maximum, and thereafter falls off as the star fades, but not as rapidly (so that, relative to the continuum, it appears to strengthen).

TABLE 2  
ABSOLUTE INTENSITY OF  $H\delta$  (LOGARITHMS OF ERGS PER SECOND)

ABSOLUTE MAGNITUDE	NOVA PERSEI	NOVA AQUILAE	NOVA CYGNI	NOVA PICTORIS	NOVA HERCULIS
-10	...	36.58	...	...	...
-9	...	36.58	...	...	...
-8	36.62	36.92	36.95	...	...
-7	36.52	36.94	36.87	35.72	35.24
-6	36.21	36.60	36.62	36.04	35.32
-5	35.80	36.42	36.25	35.44	34.68
-4	35.44	36.14	35.84	35.46	33.36
-3	35.14	35.80	35.54	35.04	34.08
-2	34.85	34.82	35.20	34.80	34.11
-1	...	...	35.00	34.36	...

The intensity of the sum of the nebular lines of  $[OIII]$ ,  $N_1$  and  $N_2$  is similarly summarized in table 3. Here the intensity remains rather steady, though from mere inspection of the spectra it appears to increase rather strongly.

TABLE 3  
ABSOLUTE INTENSITY OF THE NEBULAR LINES

ABSOLUTE MAGNITUDE	NOVA PERSEI	NOVA AQUILAE	NOVA CYGNI	NOVA PICTORIS	NOVA HERCULIS
-5	...	37.10	...	...	...
-4	...	37.28	...	...	34.48
-3	35.58	37.28	35.85	35.34	35.32
-2	35.18	37.28	35.82	35.34	35.90

3. *Physical Conditions. The Excitation Temperatures.*—The intensities of the bright lines furnish several clues to the physical condition of the material ejected. Some approximation to the problem may be obtained by treating the ejected envelope as though it were the envelope of a planetary nebula, with which a nova presents obvious analogies. The intensities of the bright lines, when compared to the corresponding intensities in the envelopes of planetaries, lead to temperatures for the exciting nuclei of novae that are very comparable to the temperatures in the central stars of the planetary nebulae. The envelope of N.G.C. 7027, for instance, can be compared with those of the novae by comparing our measured relative intensities with those published by Wyse<sup>2</sup> for that object. The excitation temperature for Nova Pictoris is found to be rather higher than for the

nebula, that of Nova Herculis, about the same and those of the other three novae, a little lower. It may be remarked that Nova Pictoris showed the highest excitation lines ever recorded for a nova, in the great intensity of [Fe VII]<sup>3</sup> in the 1931 spectrum, as described by Spencer Jones;<sup>4</sup> a high temperature is, therefore, not surprising. The temperature of the central star of N.G.C. 7027 is estimated by Aller<sup>5</sup> as 88,000°. By another method, Sayer<sup>6</sup> deduced that the excitation temperature of the nuclear star of Nova Ophiuchi 1933 (RS Ophiuchi) could not have been less than 10<sup>5</sup> degrees.

4. *Physical Conditions. The Electron Temperatures.*—The electron temperatures in the envelopes of the novae may be deduced, by the methods developed by Menzel, Aller and Hebb<sup>7</sup> for the envelopes of planetary nebulae, from the ratio of intensity of the auroral to the nebular lines of [OIII]. If the electron density is high enough, a similar method may be used for [OI]. The resulting electron temperatures are summarized in table 4.

TABLE 4  
ELECTRON TEMPERATURES FOR FIVE NOVAE

ABSOLUTE MAGNITUDE	NOVA PERSEI [OIII]	NOVA AQUILAE [OIII]	NOVA CYGNI [OIII]	NOVA PICTORIS [OIII]	NOVA HERCULIS [OIII]	NOVA HERCULIS [OI]
-7	..	..	..	..	..	6600
-6	..	..	..	..	..	6500
-5	..	5900	..	..	..	3900
-4	..	6500	..	..	7300	3500
-3	8100	5800	5800	8400	7200	..
-2	7300	5500	5900	6700	6100	..
-1	..	..	5600	6300	..	..

The electron temperatures are low, and this is readily accounted for by the depressing effect<sup>8</sup> of the abundance of OIII, and also, for several of the novae, Ne III. The very low electron temperatures deduced for Nova Herculis from the [OI] lines are without doubt due to the departure from the assumed Boltzmann distribution for this atom.

5. *Physical Conditions during the Secondary Fluctuations.*—Many novae have undergone fluctuations on the decline from maximum, the most striking cases being those of Nova Persei and Nova Aquilae. It has been shown elsewhere for Nova Aquilae<sup>9</sup> that the fluctuations are even more marked in the nuclear star than in integrated light, and these fluctuations have been interpreted<sup>10</sup> as oscillations of temperature, accompanied by apparent contractions and expansions of the central star as the temperature rises or falls. This phenomenon is in harmony with the conclusions drawn<sup>11</sup> for Nova Herculis on the hypothesis of continuous ejection: "The position of the photosphere... is governed, not by the distance to which the ejected matter has traveled, but by the physical state of the extended atmosphere and the amount of energy incident on it from beneath. The photosphere

is stationary...but matter passes through it continuously from below..." The drops in brightness, then, should represent rises of temperature and apparent contractions of the central star.

Spectroscopic evidence that this is the case is shown in figure 1, which follows the change in the ratio of He I to He II lines through typical minima in the secondary fluctuations of Nova Persei and Nova Aquilae. The ratio falls sharply during the minima, which may be ascribed to increased ionization and associated with a rise in temperature.

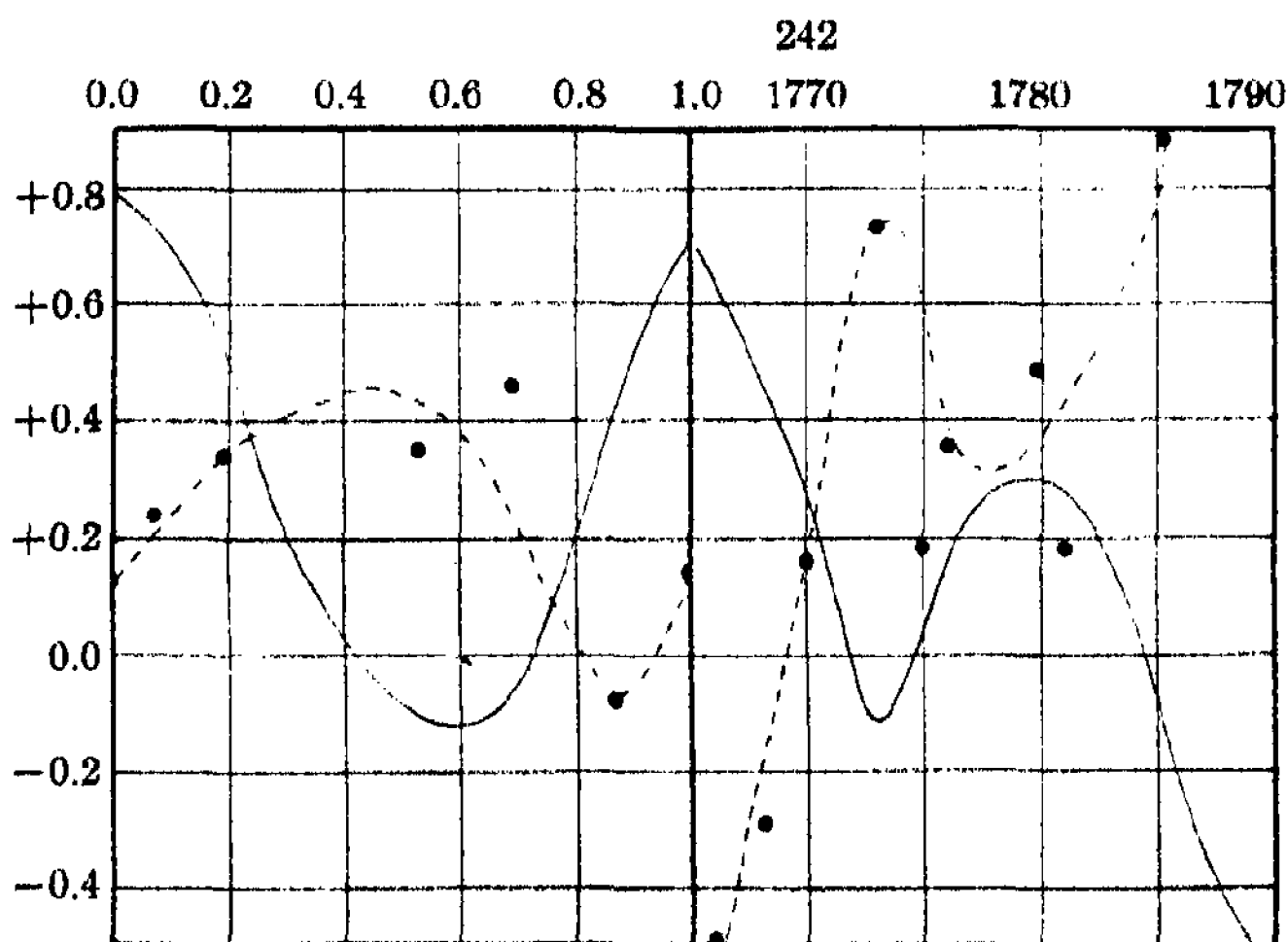


FIGURE 1

Ionization of helium during the great fluctuations of Nova Persei (left) and Nova Aquilae. Abscissae are fractions of the period of oscillation for Nova Persei, Julian days for Nova Aquilae. Ordinates are logarithms of the ratio of intensity of 4471 He I to 4686 He II. Continuous lines represent the course of the light curve, broken lines, that of the He I/He II ratio. The diagram for Nova Persei is composite; that for Nova Aquilae represents continuous observations.

6. *The Electron Densities.*—The analogy with the envelopes of planetary nebulae may be used to estimate the electron densities in the expanding atmospheres of novae. The method given by Menzel and Aller<sup>12</sup> was used, with the modifications necessary to apply to H $\delta$  (the line given in our tables), and revised values of the constants, kindly supplied by Dr. Aller.

The electron densities, which are illustrated for a number of absolute visual magnitudes for the five novae, in figure 2, are from  $10^8$  to  $10^9$  per cubic centimeter at the beginning of the outburst, and fall to between  $10^5$  and  $10^6$  per cubic centimeter during the interval covered by our measures. This

value is interesting as falling between those given by Menzel and Aller<sup>13</sup> for the envelopes of planetary nebulae ( $10^8$  to  $10^4$  per cubic centimeter) and those calculated by Aller<sup>6</sup> for the envelopes of Wolf-Rayet stars (from  $10^{12}$  to  $10^{13}$  per cubic centimeter).

7. *The Ejected Masses.*—On the very simple assumption that every electron is matched in the atmosphere by a hydrogen nucleus, the data of figure 2 may be used to estimate the masses ejected during the first few

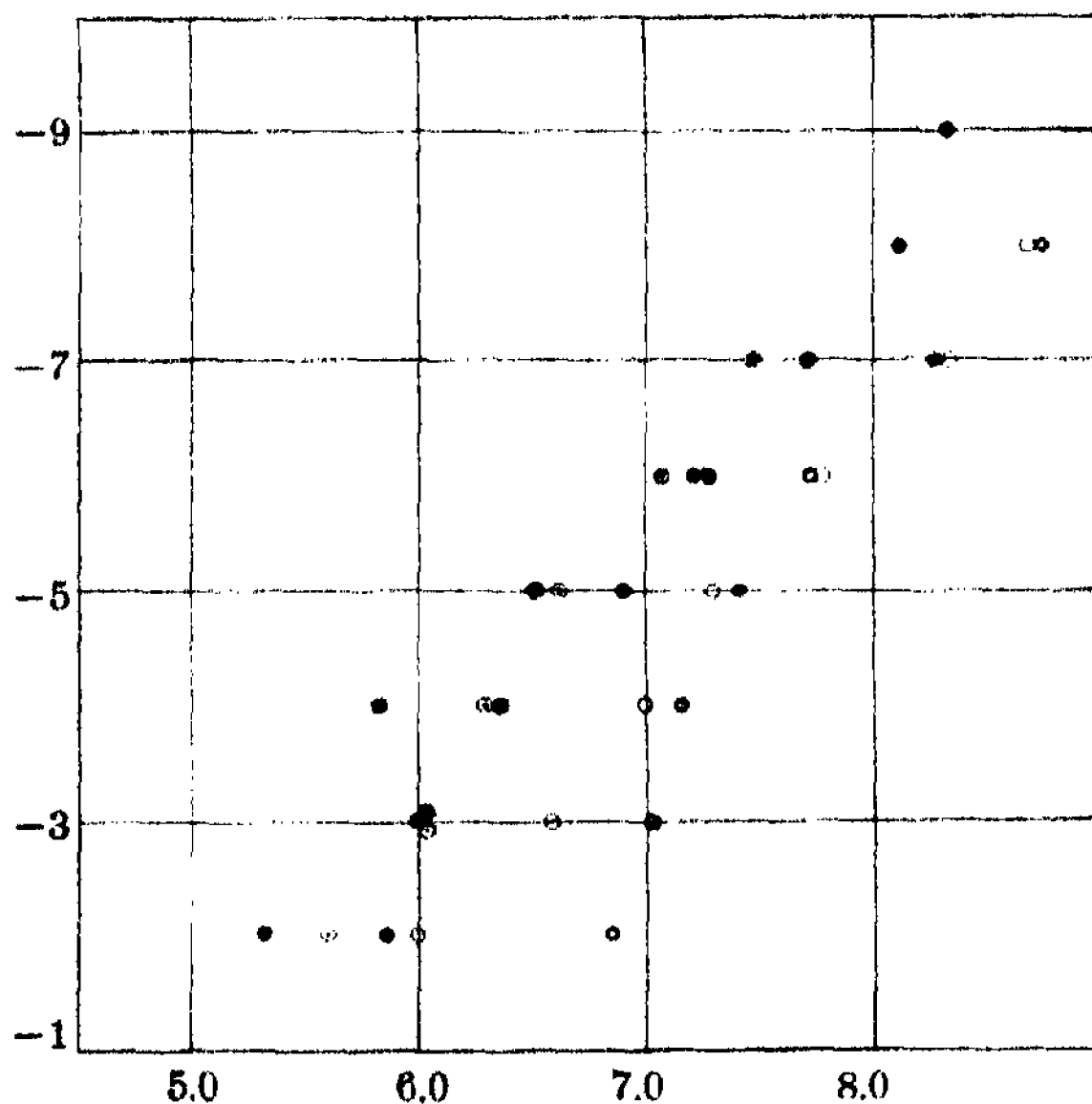


FIGURE 2

Electron densities and absolute magnitudes for five novae. Abscissae are logarithms of electron densities per cubic centimeter; ordinates are absolute visual magnitudes. The novae are distinguished as follows: Nova Persei, circles; Nova Aquilae, dots; Nova Pictoris, small dots in circles; Nova Cygni, two circles; Nova Herculis, large dots in circles.

months. The calculated masses are shown in figure 3. Evidently they continue to increase long after the time at which we must suppose ejection (in any considerable quantities) to have ceased. That this increase is real seems unlikely. More probably the hydrogen was largely ionized at the maximum, and later recombined; thus the inferred number of hydrogen atoms per cubic centimeter in the early stages (perhaps in all stages) would be too small. The assumption that equal numbers of hydrogen nuclei and electrons are present can, however, be replaced by a better one, as will be

shown in a later note. The modification alters our view concerning the number of atoms of hydrogen, but the order of the deduced masses is not changed.

It seems likely that the values deduced at the end of the period studied are nearest to the actually ejected masses. Nova Aquilae stands out as having ejected the greatest mass, just as it stands out in luminosity. Luminosity and ejected mass are compared in figure 4. The three high-weight novae (for which the actual rates of expansion of the ejected neb-

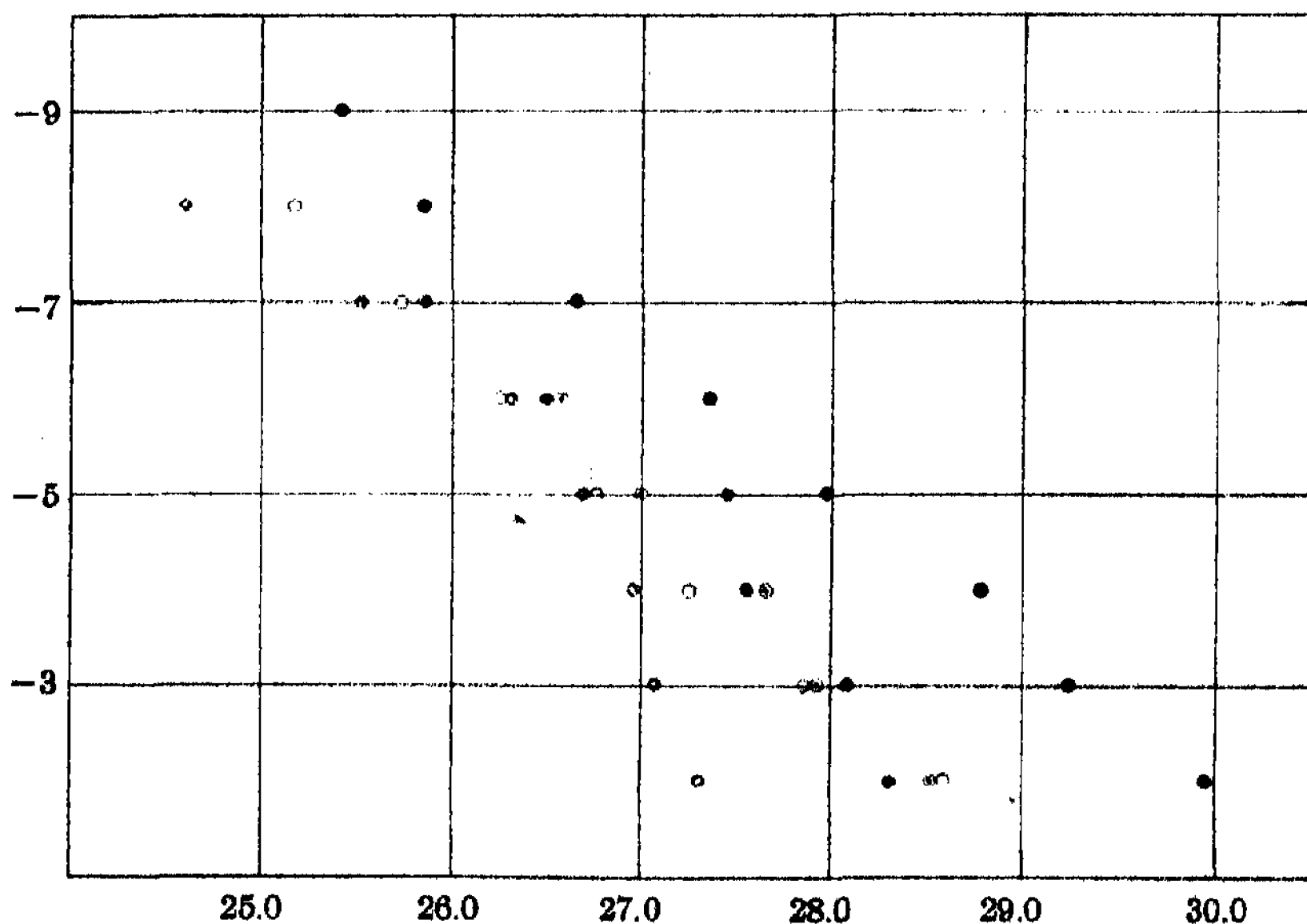


FIGURE 3

Ejected masses for five novae. Abscissae are logarithms of ejected masses, in grams. Ordinates are absolute visual magnitudes. The novae are denoted by symbols as in figure 2.

ulae are known) are shown by larger symbols in the figure. For Nova Cygni and Nova Pictoris, the size of the hypothetical expanding envelope was calculated from the radial velocities, and an arbitrary choice of velocity had to be made. A larger velocity than the 750 km./sec. used for Nova Cygni would have reduced the electron densities in the later stages, increased the deduced masses and brought it into accordance with the other novae in figure 4.

The straight line drawn in figure 4 represents proportionality between ejected material and luminosity at maximum. It is given by the relation

$$\log \text{mass} = -0.4M_{\text{vis}} + 25.5,$$

where the mass is given in grams. The three high-weight novae do not deviate much from this line.

Other estimates of the masses ejected from novae are in general agreement with the foregoing. Dr. Whipple and one of the writers<sup>14</sup> found that Nova Herculis ejected a mass of the order of  $10^{29}$  grams up to the time of the great minimum. Sayer<sup>5</sup> calculated that  $10^{28}$  grams were ejected by Nova Ophiuchi 1933. A similar estimate,  $10^{-5}$  to  $10^{-6}$  solar masses, is given<sup>15</sup> by Ambarzumian and Gordeladze. For Nova Lacertae 1936, Popper<sup>16</sup> has derived masses of  $6 \times 10^{-5}$  and  $7 \times 10^{-5}$  solar masses, from electron densities of the order of  $10^6$  per cubic centimeter about 100 days after maximum light.

It is hoped to study the composition of the ejected material in a later communication.

8. *Summary.*—The present note draws some elementary physical conclusions from the current measures of the intensities of the bright lines in the spectra of five novae. The excitation temperatures are very high—comparable to those in the nuclei of planetary nebulae. The electron temperatures are less than  $10,000^\circ$ , not very different from those in the envelopes of planetary nebulae. The great fluctuations of Nova Persei and Nova Aquilae were occasioned by sharp rises and falls of excitation temperature.

The electron densities, estimated by analogy with the envelopes of planetary nebulae, are from  $10^8$  to  $10^9$  per cubic centimeter at the outburst, and fall (during the interval studied) to  $10^5$  or  $10^6$  per cubic centimeter. Thus the expanding envelopes of novae are intermediate between the planetary nebulae and the envelopes of Wolf-Rayet stars.

The mass of ejected material is estimated at from  $10^{27}$  to  $10^{30}$  grams for the five novae studied. These masses appear to be in rough proportion to the luminosities of the corresponding novae at maximum.

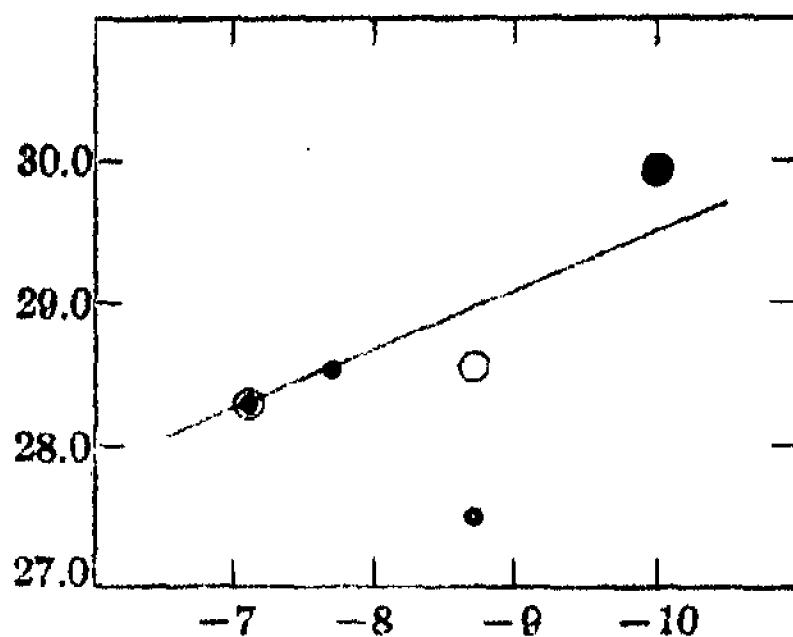


FIGURE 4

Relation of ejected mass to luminosity at maximum. Ordinates are logarithms of ejected masses in grams. Abscissae are absolute visual magnitudes at maximum. The five novae are denoted by the same symbols as in figures 2 and 3. Points of lower weight are shown by small symbols. The straight line represents the relation:  $\log \text{mass} = -0.4M_{\text{vis}} + 25.5$ .

<sup>14</sup> Whipple, F. L., and Payne-Gaposchkin, C., *Harv. Obs. Circ.*, No. 414, 1-13 (1937); Payne-Gaposchkin, C., and Whipple, F. L., *Ibid.*, No. 433, 1-12 (1939); Payne-

Gaposchkin, C., and Menzel, D. H., *Ibid.*, No. 428, 1-13 (1936); Payne-Gaposchkin, C., and Gaposchkin, S., *Ibid.*, No. 445, 1-19 (1942).

<sup>2</sup> Wyse, A. H., *Astrophys. Jour.*, **95**, 356-385 (1942).

<sup>3</sup> Bowen, I. S., and Wyse, A. H., *Lick Obs. Bull.*, **19**, 1-16 (1939); Bowen, I. S., and Edlén, B., *Nature*, **143**, 374 (1939).

<sup>4</sup> Spencer Jones, H., *Cape Obs. Ann.*, **10**, Part 9, 144 (1931).

<sup>5</sup> Private communication.

<sup>6</sup> Thesis, Harvard (1935).

<sup>7</sup> Menzel, D. H., Aller, L. H., and Hebb, M. H., *Astrophys. Jour.*, **93**, 230-235 (1941).

<sup>8</sup> Menzel, D. H., and Aller, L. H., *Ibid.*, **94**, 30-36 (1941).

<sup>9</sup> Payne-Gaposchkin, C., *Les Novae et les Naines Blanches*, Paris (1939).

<sup>10</sup> Payne-Gaposchkin, C., *Ann. Astrophys. Obs.* (Tonanzintla, Mex.), **1** (1942) (in press).

<sup>11</sup> Whipple, F. L., and Payne-Gaposchkin, C., *Harv. Obs. Circ.*, No. 413, 1-12 (1936).

<sup>12</sup> Menzel, D. H., and Aller, L. H., *Astrophys. Jour.*, **93**, 195-201 (1941).

<sup>13</sup> Menzel, D. H., and Aller, L. H., *Ibid.*, **93**, p. 199 (1941).

<sup>14</sup> Whipple, F. L., and Payne-Gaposchkin, C., *Harv. Obs. Circ.*, No. 413, p. 10 (1936).

<sup>15</sup> Krat, W., *Pulk. Obs. Circ.* No. 29, 21-25 (1940).

<sup>16</sup> Popper, D. M., *Publ. Astr. Soc. Pacific*, **53**, 252-255 (1941).

## ON THE DIMENSIONS AND CONSTITUTION OF VARIABLE STARS

BY CECILIA PAYNE-GAPOSCHKIN AND SERGEI GAPOSCHKIN

HARVARD COLLEGE OBSERVATORY

Communicated October 16, 1942

1. *The Observational Material.*—A unified study of the properties of variable stars has been carried out by the writers during the past few years.<sup>1</sup> With the accumulation of material relative to the periods, magnitudes and spectra of variable stars it is becoming possible not only to examine them as individuals, but to group them in families with common physical characteristics, and thus, perhaps, to approach the problems of their development and origin.

A general summary of what is now known or inferred concerning the principal physical parameters that describe the intrinsic variables—their luminosities, radii and masses—has been presented elsewhere by the authors.<sup>2</sup> References to the sources of the material are given in the paper cited. The principal results are shown graphically in figures 1 and 2. Figure 1 shows, on the left, the relation to the logarithm of the luminosity (in terms of the sun) to the logarithm of the period for Cepheid variables; on the right, the relationship of spectrum to logarithm of luminosity for the red variables, both semi-regular and long-period.

Figure 2 relates the sizes of Cepheid variables to their periods, and the

sizes of red variables to their spectral classes. The unit of size is the radius of the sun.

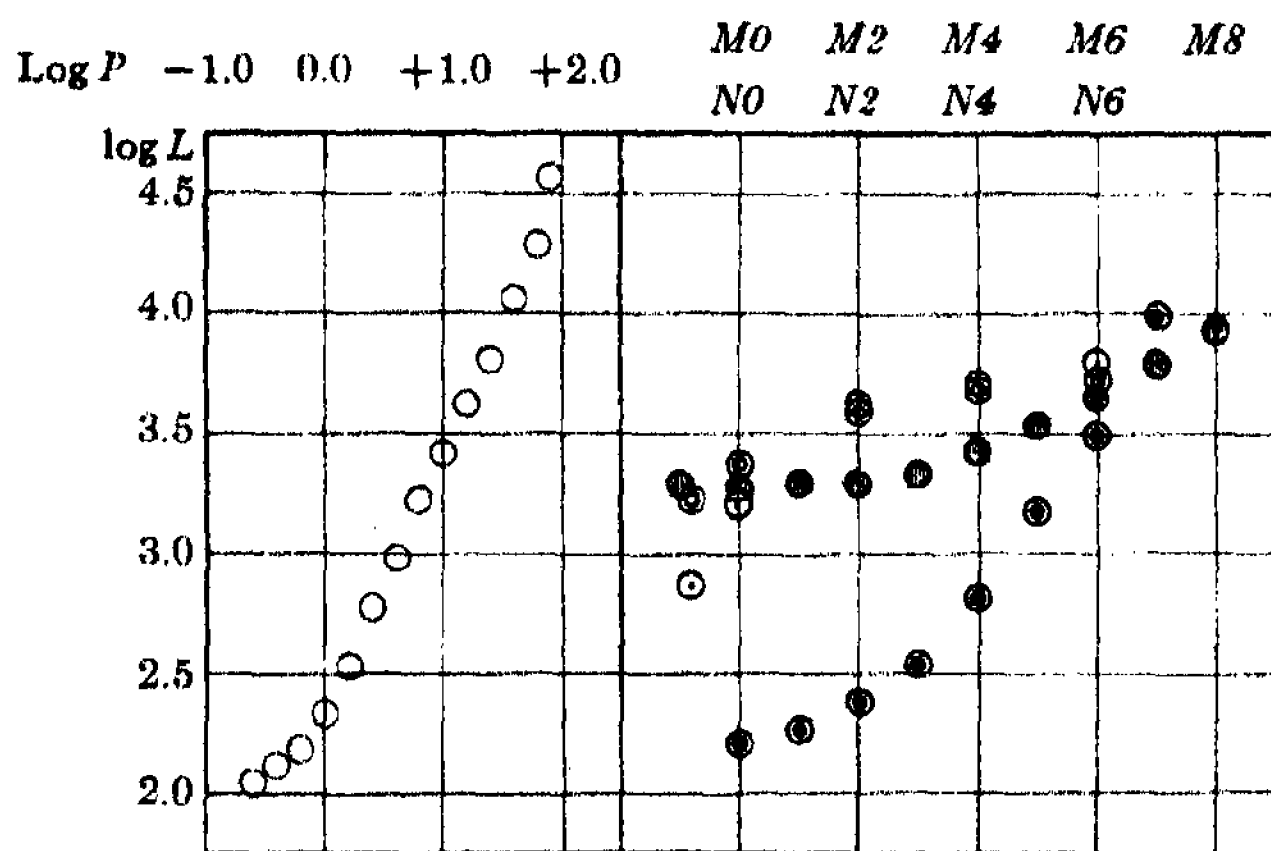


FIGURE 1

Luminosities of intrinsic variables as functions of period or spectrum. Ordinates are logarithms of luminosities in terms of the sun. The various types of variables are denoted by symbols as follows: Cepheid variables, circles; long-period *M* stars, shaded dots; long-period *N* stars, circles in circles; semi-regular *M* stars, large dots in circles; semi-regular *N* stars, small dots in circles.

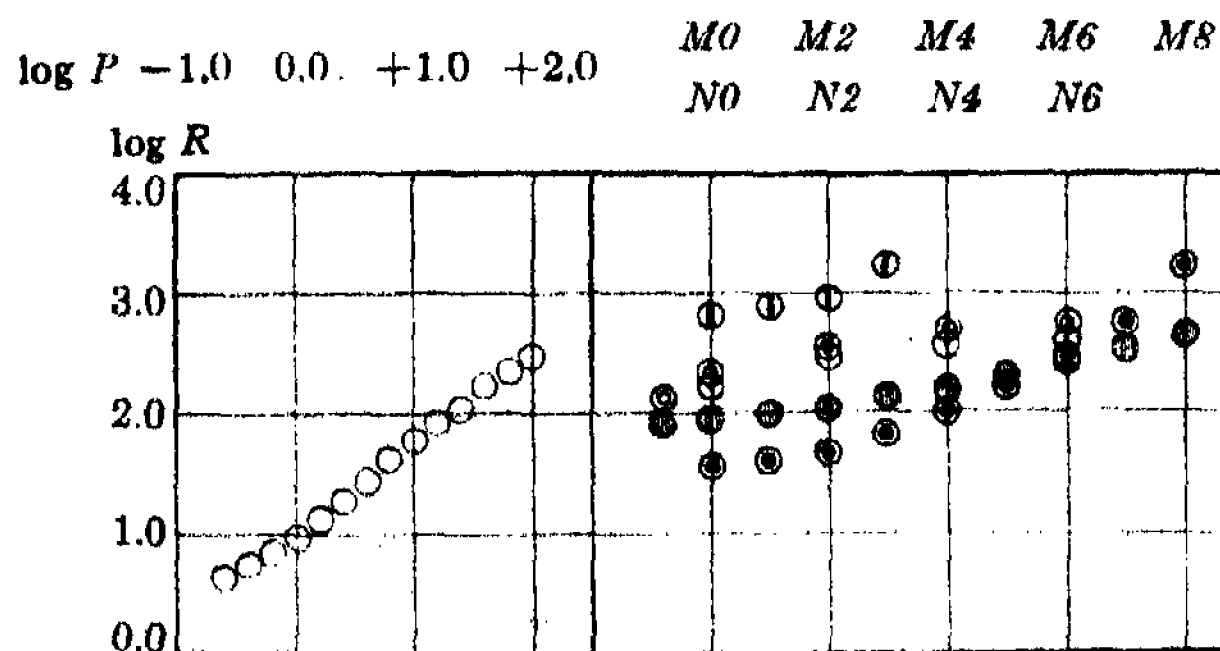


FIGURE 2

Sizes of intrinsic variables. Abscissae are logarithms of period (left) for Cepheids, spectra, for long-period and semi-regular variables. Ordinates are logarithms of radii, in terms of the solar radius. Symbols are as in figure 1. Barred circles refer to supergiant semi-regular variables.

It is seen from figures 1 and 2 that in luminosity and in size the Cepheids and the red variables greatly exceed the sun. Neither in size nor in luminosity, however, do the two groups constitute a continuous series. Per-



haps the supergiant red variables in figure 2 may be considered to follow from the Cepheids in dimensions.

2. *The Constitution of Variable Stars.*—Our knowledge of the internal conditions of the stars depends upon a knowledge of luminosity, mass, radius and molecular weight. In the calculations that will now be quoted, the mean molecular weight has been assumed to be 1.00, corresponding to a hydrogen content of 0.33. The central condensation (and, consequently, the polytropic index) has been calculated from the data of figures 1 and 2 by the methods of Chandrasekhar.<sup>3</sup> The mean polytropic indices thus obtained for the different groups of variables are as follows: Cepheid variables, 2.96; long-period variables, Class *M*, 3.15; long-period variables, Class *N*, 3.30; semi-regular variables, Class *M*, 3.19; semi-regular variables, Class *N*, 3.30, supergiant (semi-regular) *M* variables, 3.8: .

There is no evidence of high central condensation, nor do any of the intrinsic variables seem to deviate from polytropic index 3 in the direction of homogeneity. The calculations have assumed that the stars follow the (empirical) mass-luminosity relation. If they are under-luminous, the polytropic indices are smaller; if they are over-luminous, the polytropic indices are larger. We have, however, no evidence on departures of intrinsic variables from the mass-luminosity relation.

A second method of estimating the constitution of intrinsic variables arises from an application of the pulsation theory. It consists essentially in finding how the polytropic index must be adjusted to give the observed relationship between period and mean density. An estimate of the ratio of the specific heats is required for the application of this method: the value 1.429 was adopted. The average results are as follows: for Cepheids, 2.47; long-period *M* stars, 2.12; long-period *N* stars, 2.30; semi-regular *M* stars, 2.54; semi-regular *N* stars, 2.98.

The above estimates of the internal conditions suggest that the point-source model<sup>4</sup> (for which the exterior corresponds to polytropic index 3.25, the interior, to 1.5) will furnish a useful approximation for the examination of internal conditions.

3. *The Central Temperatures.*—For a star built on the point-source model, the central temperature  $T_c$  is given by the formula:<sup>5</sup>

$$\log T_c = 7.3175 + \log M - \log R,$$

where  $M$  and  $R$  are mass and radius in terms of those of the sun. For the standard model the temperature would be given by the formula:<sup>6</sup>

$$\log T_c = 7.2949 + \log \mu + \log \beta + \log M - \log R,$$

where  $\mu$  and  $\beta$  are mean molecular weight (assumed equal to 1.00) and the ratio of radiation pressure to total pressure (appreciably less than unity for these high masses). The effect of the constant and the factor  $\beta$  is to

make the standard-model temperatures rather lower than the point-source model temperatures. The latter are indeed the *highest* that are compatible with the observational data.

Values of  $(\log M - \log R)$ , and the corresponding temperatures (point-source model) are given in table 1. The relation of these central temperatures to period and spectrum for the variable stars concerned is shown in figure 3.

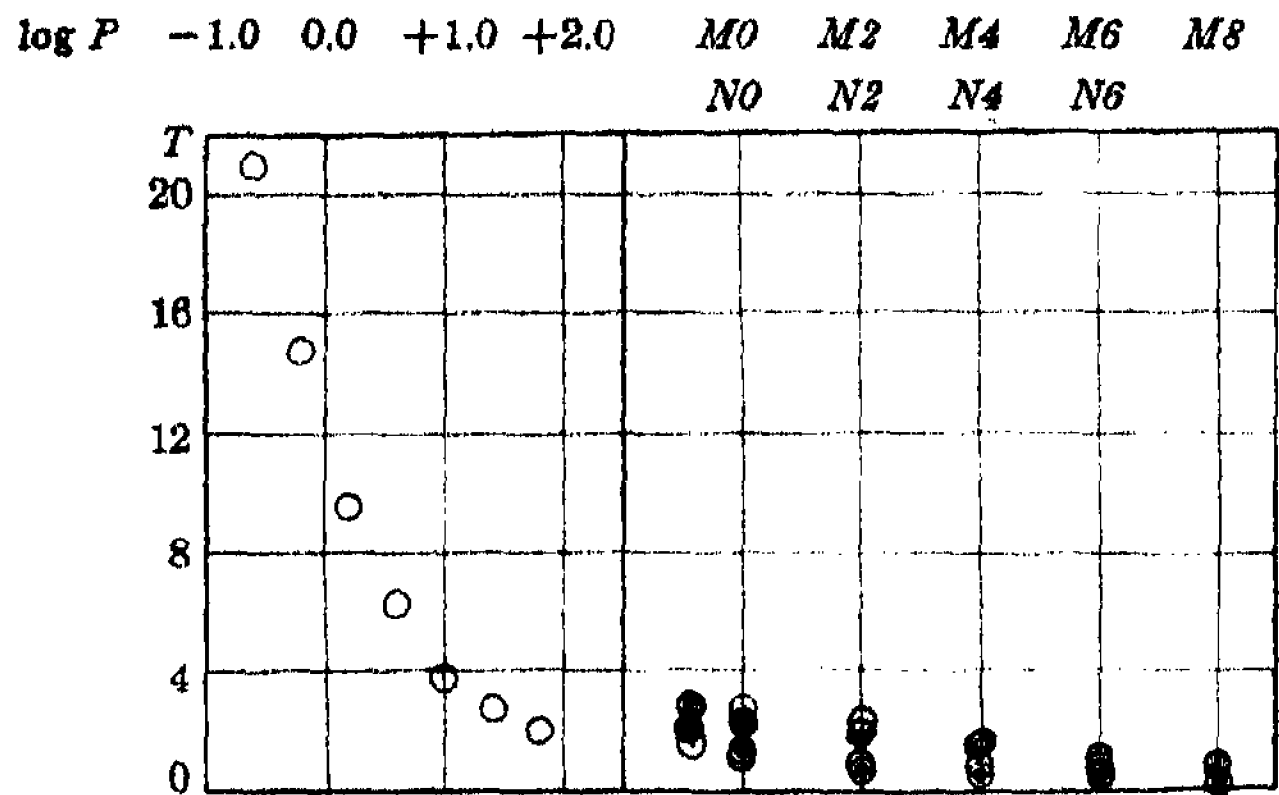


FIGURE 3

Central temperatures for different types of variable stars. Abscissae are as in figures 1 and 2. Ordinates are central temperatures in millions of degrees centigrade. Symbols are as in figures 1 and 2.

TABLE 1

CENTRAL TEMPERATURES FOR VARIABLE STARS

CEPHEID VARIABLES			LONG-PERIOD VARIABLES			SEMI-REGULAR VARIABLES		
LOG P	LOG M/R	LOG T <sub>c</sub>	SPECTRUM	LOG M/R	LOG T <sub>c</sub>	SPECTRUM	LOG M/R	LOG T <sub>c</sub>
-0.6	0.00	7.82	K5e	-0.92	6.40	M0	-0.87	6.45
-0.2	-0.15	7.17	M0e	-0.95	6.37	M2	-0.95	6.37
0.2	-0.34	6.98	M2e	-1.02	6.30	M4	-1.12	6.20
0.4	-0.43	6.89	M4e	-1.13	6.19	M6	-1.37	5.95
0.6	-0.52	6.80	M6e	-1.30	6.02	M8	-1.74	5.58
0.8	-0.64	6.68	M8e	-1.43	5.89	M9	-1.93	5.39
1.0	-0.73	6.59	....	....	....	...	....	....
1.2	-0.79	6.53	R8e	-1.12	6.20	R8	-1.14	6.18
1.4	-0.88	6.44	NOe	-1.29	6.03	NO	-1.25	6.07
1.6	-0.96	6.36	N2e	-1.45	5.87	N2	-1.37	5.95
1.8	-1.00	6.32	N4e	-1.55	5.77	N4	-1.41	5.91
2.0	-1.06	6.26	N6e	-1.60	5.72	N6	-1.42	5.90

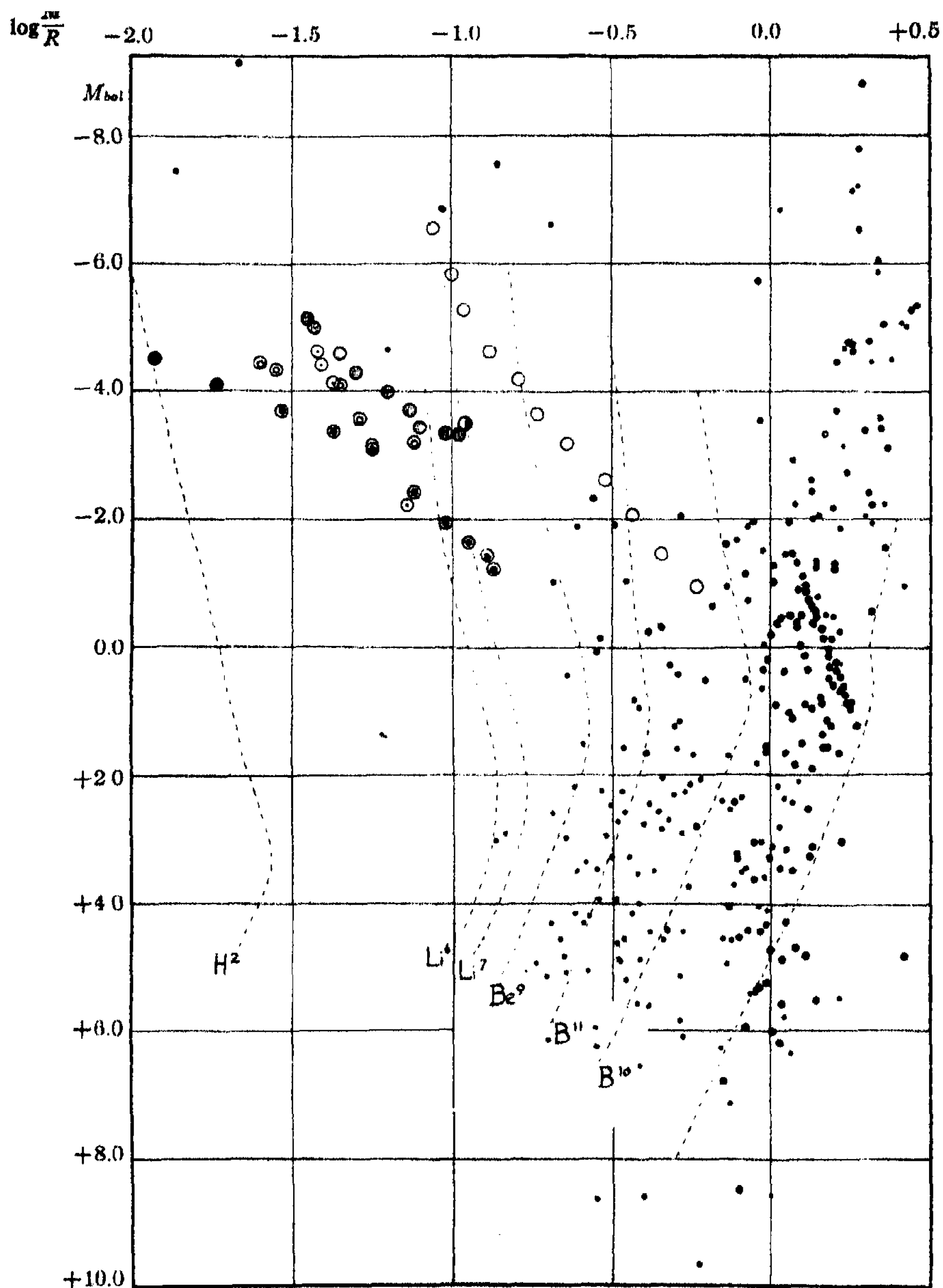


FIGURE 4

Relation between absolute bolometric magnitude (ordinate) and  $\log M/R$  in solar units. The five groups of variable stars are denoted by symbols as in figures 1 to 3. The half-filled circle refers to RV Tauri stars. Smaller dots denote the first components of selected eclipsing variables; the smallest dots refer to the second components. Broken lines indicate the course of the subatomic reactions.

4. *The Subatomic Processes.*—The relationship, between central temperature of stars and the atomic processes that are the source of the radiated stellar energy, has been extensively discussed in recent years.<sup>7</sup> That the red variable stars probably utilize lighter elements than those used in the carbon cycle has been shown by Gamow and his associates.<sup>8</sup> The rôle played by the various subatomic processes is illustrated in figure 4, which presents the relationship between  $\log M/R$  and absolute bolometric magnitude for the families of variable stars, and also for a number of other stars of well-known dimensions. Lines corresponding to the various subatomic processes are inserted in the figure.

Several suggestive conclusions can be drawn from figure 4. No stars fall in the region of deuterium as an energy source. Most of the long-period and semi-regular red variables fall in the lithium region. The Cepheids span the beryllium and boron regions, with a suggestion of transition from one subatomic process to another near the minima of period frequency. The point representing the RV Tauri stars<sup>9</sup> places them in such a region of transition, near the limit of the long-period  $M$  stars. The bright components of eclipsing stars provide a specimen of "normal" stars that fall in the region of the carbon cycle.

There are, however, a number of double star systems of which the two members fall in very different regions of the figure. That two members of one system are drawing on different energy sources is by no means a rare phenomenon, and raises provocative questions concerning the origin and evolution of such stellar pairs.

<sup>1</sup> Payne-Gaposchkin, C., *Amer. Phil. Soc. Proc.*, **81**, 189 (1939).

<sup>2</sup> Payne-Gaposchkin, C., and Gaposchkin, S., *Amer. Phil. Soc. Proc.* (1942) (in press.)

<sup>3</sup> Chandrasekhar, S., *Mon. Not., Roy. Astr. Soc.*, **96**, 647-660 (1936).

<sup>4</sup> Chandrasekhar, S., *An Introduction to the Study of Stellar Structure*, Chicago, p. 349 ff. (1938).

<sup>5</sup> Chandrasekhar, S., *Ibid.*, p. 354 (1939).

<sup>6</sup> Chandrasekhar, S., *Ibid.*, p. 230 (1939).

<sup>7</sup> Bethe, H., *Phys. Rev.*, **55**, 103 (1938); *Ibid.*, **55**, 434-456 (1939).

<sup>8</sup> Gamow, G., *Ibid.*, **55**, 718-725 (1939); Gamow, G., and Teller, E., *Ibid.*, **55**, 791 (1939); Greenfield, M., *Ibid.*, **60**, 175-183 (1941).

<sup>9</sup> Payne-Gaposchkin, C., Brenton, V. K., and Gaposchkin, S., *Harv. Obs. Ann.*, **113**, No. 1 (1942) (in press); Payne-Gaposchkin, C., and Brenton, V. K., these PROCEEDINGS, **28**, 496-500 (1942).

## A STUDY OF THE RV TAURI VARIABLES

BY CECILIA PAYNE-GAPOSCHKIN AND VIRGINIA K. BRENTON

HARVARD COLLEGE OBSERVATORY

Communicated October 16, 1942

1. *Definitions.*—Among the intrinsic variables, the RV Tauri stars stand out with rather definite characteristics. Three principal criteria may be used for assigning stars to the class.

(a) Successive minima differ in depth. Deep and shallow minima are conventionally called "primary" and "secondary."

(b) Primary and secondary minima interchange phases at intervals.

(c) The stars are variable in spectrum and in radial velocity. The spectrum is of Class *F*, *G* or *K* at maximum.

The shortest known period is 38 days (EZ Aquilae), the longest is 146 days (R Scuti). Although the period usually assigned to these stars is that from one primary minimum to the next, a comparison with the velocity curve shows that the star undergoes a type of pulsation with one-half that period.

The brighter RV Tauri stars have recently been studied on the Harvard collection of stellar photographs. Of the twenty-nine stars now assigned to the class, fourteen proved to be bright enough for intensive study. A discussion of about sixteen thousand observations of these stars has been completed,<sup>1</sup> and the present note is intended to summarize some of the physical results.

2. *The Secondary Characteristics.*—Many RV Tauri stars show rather marked irregularities in period, and though over a long interval a mean period can always be assigned, instantaneous elements for short intervals may require periods differing by one or two per cent. This characteristic is not peculiar to RV Tauri stars: it is shared by other semi-regular variables and by long-period variables.

Many RV Tauri stars also show fluctuations of mean brightness; very marked for stars like DF Cygni and somewhat marked for AR Puppis (for both of which the fluctuation is periodic), marked, but not periodic for R Sagittae, and less conspicuous for U Monocerotis and R Scuti. On the whole, the fluctuations of mean brightness seem most marked for stars whose double period is near 75 days. For RV Tauri stars of longest period they are small. It may be noted on the other hand that interchanges of minima seem to be commonest for the stars of longest period; for stars like DF Cygni and AC Herculis they occur very rarely. Fluctuations of mean brightness are found for the AI Scorpii and  $\mu$  Cephei stars, as well as for RV Tauri stars.

From comparison of the visual light curve with the photographic one, it

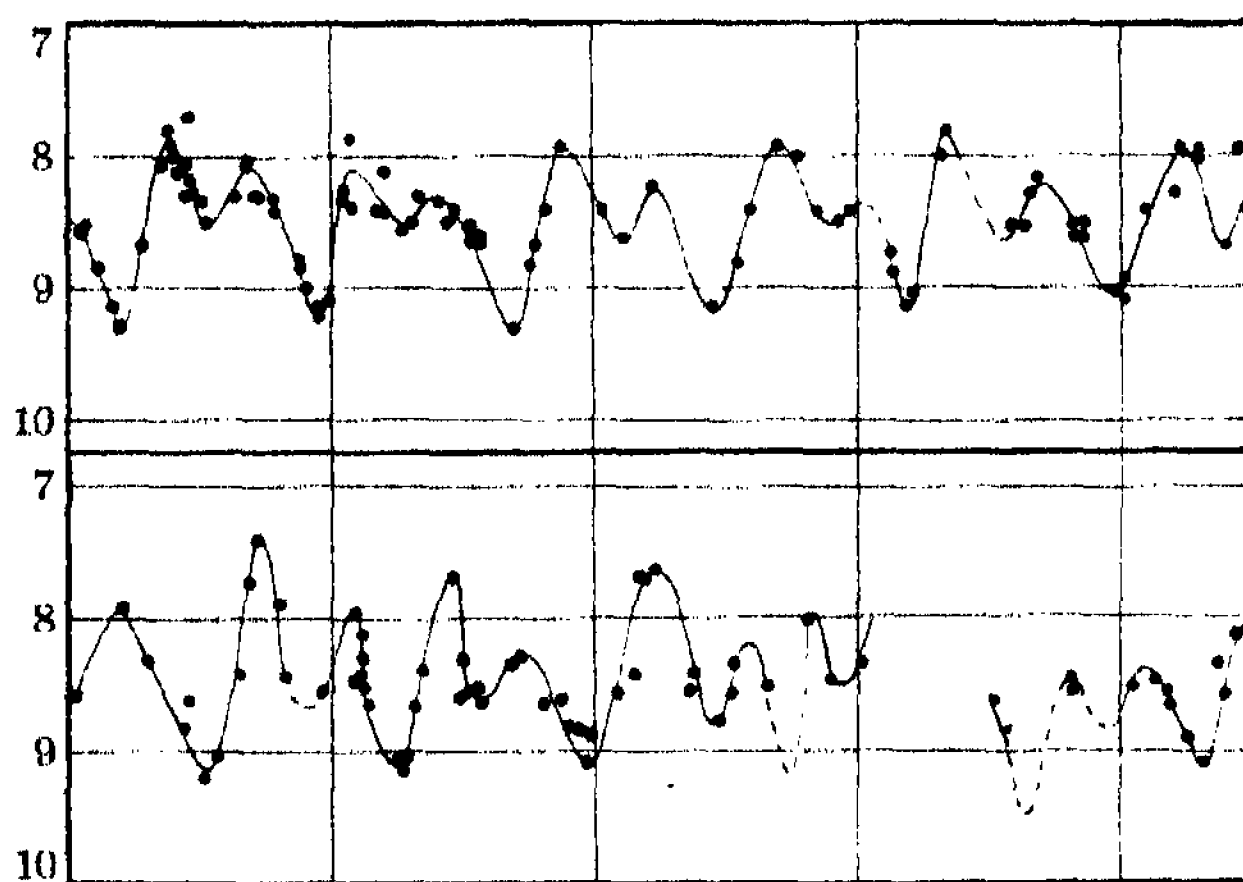


FIGURE 1

Variations of AC Herculis, a typical regular RV Tauri star. Ordinates are photographic magnitudes. Vertical lines are drawn at intervals of 100 days.

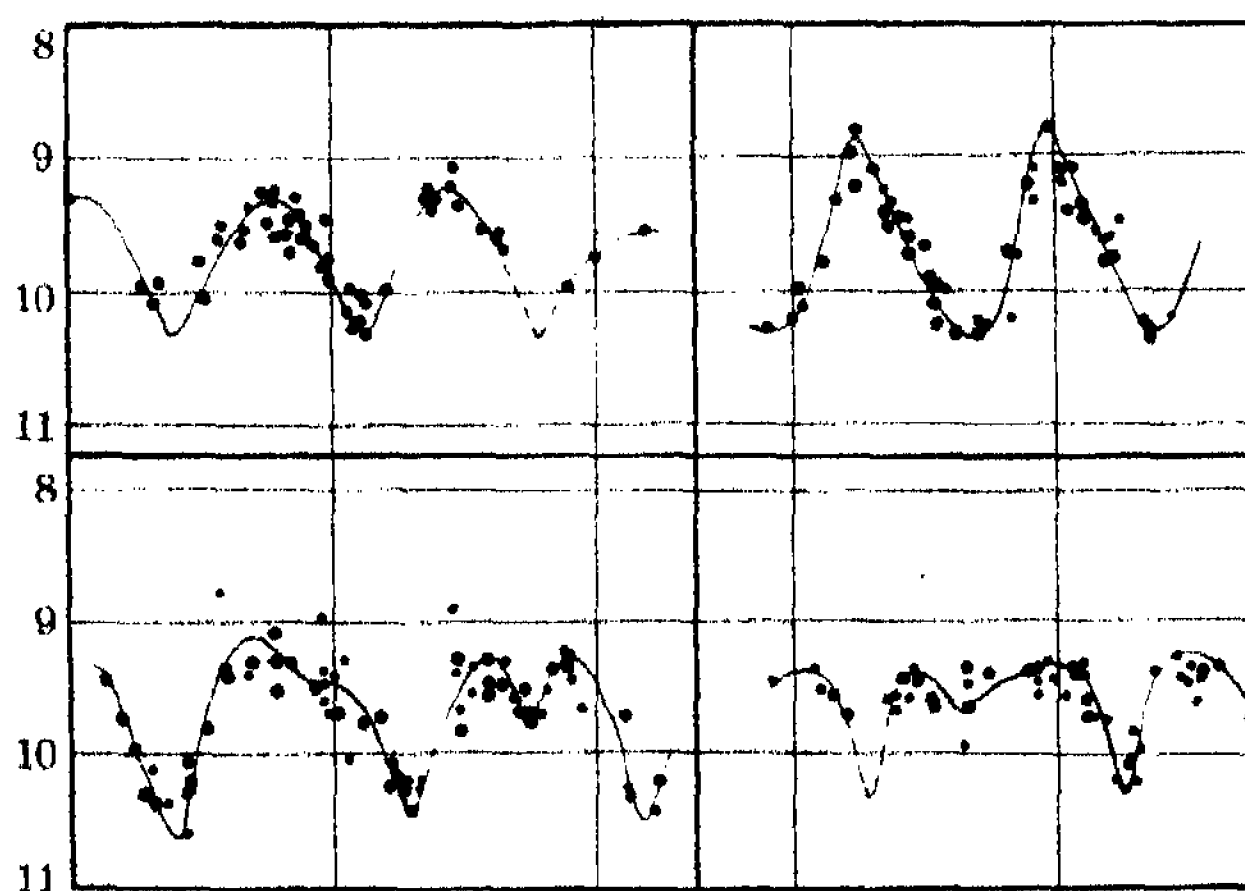


FIGURE 2

Variations of UU Herculis. Ordinates are photographic magnitudes. Vertical lines are drawn at intervals of 100 days. Above, the 72-day "fundamental" light curve. Below, the 45-day "overtone" light curve.

is evident that minima, both primary and secondary, occur somewhat earlier photographically. A rough measure of the difference in time indicates that on an average the photographic curve precedes the visual by

0<sup>p</sup>.026. This value is twice as large as the shift that Kukarkin<sup>2</sup> found for the long-period Cepheids.

3. *The Variations of UU Herculis*.—Among the twenty-nine stars assigned to the RV Tauri group,<sup>1</sup> UU Herculis shows the characteristics only intermittently. The peculiarities of this star were pointed out by Gerasimovič some years ago.<sup>3</sup> Subsequent work has substantiated his conclusion that the star alternates between two periods: an RV Tauri-like curve shows a period of about 90 days, and a Cepheid-like curve, a period of 72 days. Three complete alternations between these two periods have been observed since 1900.

Figure 1 shows some excerpts from the Harvard photographic light curve of AC Herculis, perhaps the most regular of the RV Tauri stars. It shows the typical aspect of the light curve. Figure 2 shows some excerpts from the variations of UU Herculis: above is the Cepheid-like light curve; below, the RV Tauri-like curve. The ratio between the physically significant half-period of the RV Tauri-like curve and the period of the Cepheid-like curve is 0.63. Both periods can be represented by linear elements throughout the observed interval.

Gerasimovič expressed the conviction that the two periods, though not in an obvious, simple ratio, could not be physically independent. More recently it has been shown by Schwarzschild<sup>4</sup> that both overtone and fundamental pulsations may be found among the pulsating stars of short period. He has calculated that the ratio between the fundamental period and the period of the first overtone should be between 0.55 and 0.74, according as the ratio of specific heats of the material within the star is between 1.43 and 1.67. It is tempting to conclude that the Cepheid-like light curve represents the fundamental for UU Herculis, while the RV Tauri-like curve (with, of course, half the conventional period) represents the overtone pulsation. The ratio between the two periods, 0.63, falls in the middle of the range given by Schwarzschild. If the RV Tauri-like variations of these stars with alternating minima are in fact overtone pulsations, the conventional period must be multiplied by about 7/20 to obtain the fundamental period that is physically comparable to the period of a Cepheid variable.

4. *Relationships between RV Tauri Stars and Other Variables*.—The periods of RV Tauri stars fall between those of Cepheids and long-period variables, with a certain amount of overlapping at both ends. They have usually been regarded as transition objects whose irregularities are related to a minimum of period frequency.

The relationship between the period and spectra of the RV Tauri stars, and the comparison with other intrinsic variables give a picture of the rôle that they play among the variable stars. The material is shown in figure 3, for which the data have been taken from the study cited earlier.<sup>1</sup> Logarithms of period are plotted against spectrum, and in addition to RV

Tauri stars, Cepheids, semi-regular yellow and red variables, and long-period variables are shown.

A striking feature of figure 3 is the large differences of spectrum for RV Tauri stars of given period. The spectra of these stars are notoriously

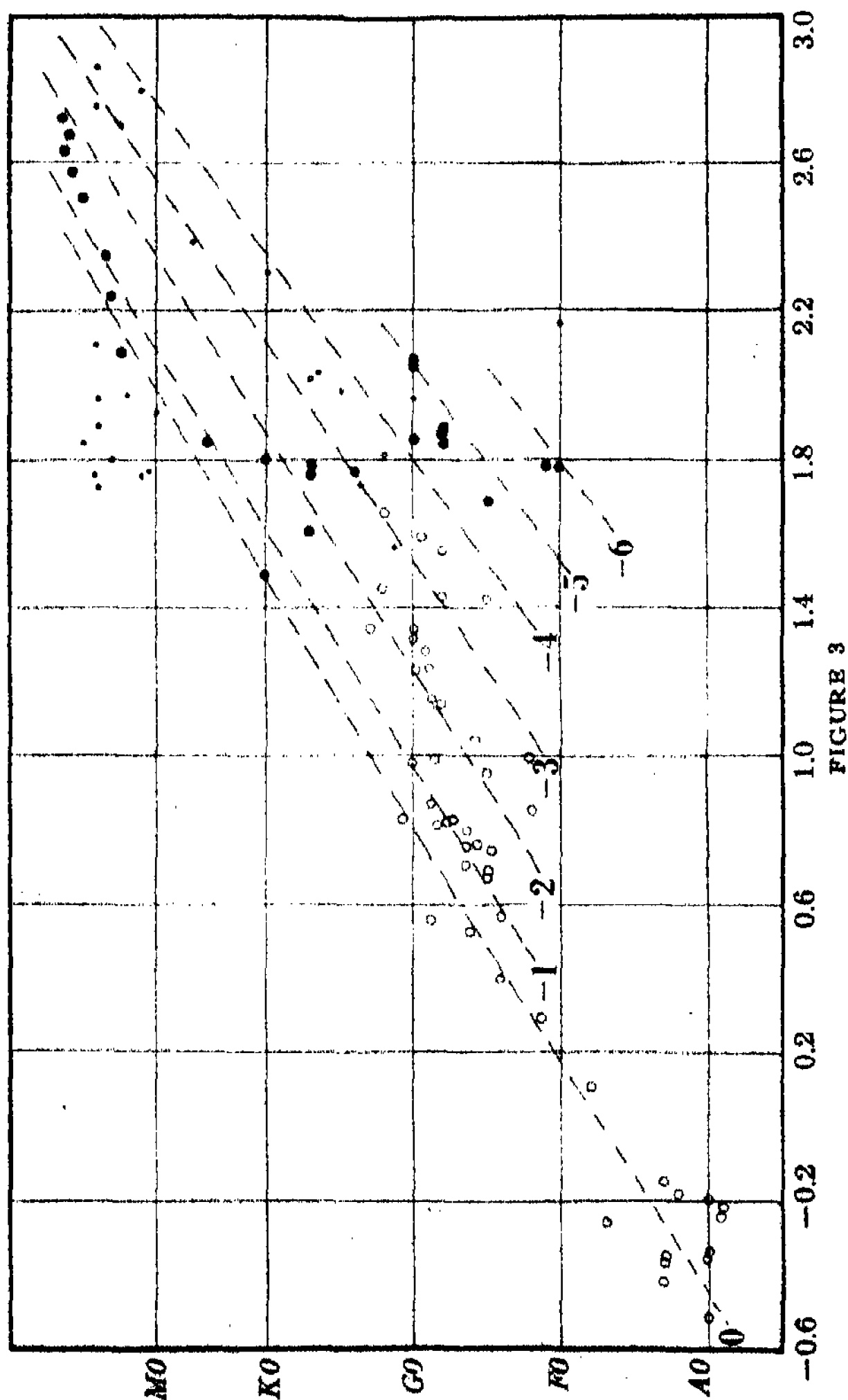


FIGURE 3

Logarithm of the period (abscissa) and maximal spectrum for intrinsic variable stars. Large dots denote RV Tauri stars, large circles, Cepheid variables; small circles, semi-regular yellow variables; small dots, semi-regular red variables; dots in circles, long-period variables (average period-spectrum relation). Broken lines denote lines of equal absolute photographic magnitude.

difficult to classify, but that AC Herculis and AR Puppis are early F stars, and DF Cygni and V Vulpeculae late G stars, are well-substantiated observations. A smaller, but quite definite, dispersion in the period-spectrum relation for Cepheids has been well known for some time.



A number of lines have been inserted in figure 3 to show the probable course of lines of equal absolute photographic magnitude. The lines have been defined at one end by means of the period-luminosity relation and the period-spectrum relation<sup>5</sup> (relating to maximum light,<sup>6</sup> for the median spectra of RV Tauri stars are deceptive). At the upper end they have been drawn with the aid of Keenan's study<sup>7</sup> of the luminosities of the small-range red variables.

If the course of the lines of equal absolute photographic magnitude be accepted, it is seen that the period-luminosity relation for the RV Tauri stars is so widely scattered as to be non-existent, though the *mean* relation would be nearly continuous with that for the Cepheids. The latter, and the semi-regular yellow variables that are not RV Tauri stars (such as AG Aurigae and SX Herculis) seem to pass through the center of the RV Tauri stars, which now appear as Cepheid-like stars of exceptional color for their periods. Very possibly the class of RV Tauri stars, marked by light curves of exceptional form, is not a physically homogeneous one.

Two stars that are certainly allied to, if not members of, the RV Tauri group, are inserted in figure 3 with crosses. These are the semi-regular variables in Messier 4<sup>8</sup> and in Omega Centauri.<sup>9</sup> They have been plotted with respect to *absolute photographic magnitude* and period. Clearly they fit closely among the RV Tauri stars, and make it seem probable that the lines of equal absolute magnitude are drawn in a roughly correct manner.

Figure 3 throws a little light upon the period-luminosity relationship for Cepheid variables. If the course of the lines in the figure is correct, it would follow that at a given period there is a dispersion of  $\pm 0.5$  in absolute photographic magnitude for Cepheids. This value may be compared with the observed scatter of the Magellanic period-luminosity relation,<sup>10</sup> which very likely refers to a more homogeneous group of stars than the galactic specimens collected in figure 3.

The photographic observations of the RV Tauri stars discussed in this paper were made possible by a grant from the Milton Fund of Harvard University.

<sup>1</sup> Payne-Gaposchkin, C., Brenton, V. K., and Gaposchkin, S., *Harv. Obs. Ann.*, 113 No. 1, 1942 (in press.)

<sup>2</sup> Kukarkin, B. W., *Nijni Novgorod Var. Stars, Nijni Novgorod*, 5, 42-45 (1936).

<sup>3</sup> Gerasimović, B. P., *Harv. Obs. Bul.*, No. 857, 27-32 (1928).

<sup>4</sup> Schwarzschild, M., *Astrophys. Jour.*, 94, 245-252 (1941).

<sup>5</sup> Shapley, H., these PROCEEDINGS, 26, 541-548 (1940).

<sup>6</sup> Payne-Gaposchkin, C., and Gaposchkin, S., *Variable Stars*, Cambridge, p. 155, (1938).

<sup>7</sup> Keenan, P. C., *Astrophys. Jour.*, 95, 461-467 (1942).

<sup>8</sup> Sawyer, H. B., *Jour. Roy. Astr. Soc. Can.*, 36, 213-217 (1942).

<sup>9</sup> Martin, W. C., *Bul. Astr. Netherlands*, 8, 290-292 (1938).

<sup>10</sup> Shapley, H., McKibben, V., and Craig, R. A., these PROCEEDINGS, 28, 192-199 (1942).

*GALACTIC AND EXTRAGALACTIC STUDIES, XVI. PHOTOGRAPHIC AMPLITUDES OF CLASSICAL CEPHEIDS*

BY HARLOW SHAPLEY

HARVARD COLLEGE OBSERVATORY

Communicated October 13, 1942

1. The correlation of the amplitude of variation with period length and with absolute magnitude for classical Cepheid variables is derived with uncertainty from existing data for the galactic Cepheids because of the following difficulties:

(a) The systems of photographic magnitudes used by the various workers depend on different telescopes and photographic emulsions; they have not been and cannot well be brought to a single uniform standard;

(b) The comparison star sequences vary in security;

(c) The colors of Cepheids are too poorly known to permit dependable reduction of visual light curves and amplitudes to the photographic system;

(d) There may be an uneven but distinct selection of stars of large range by some observers.

For 86 variables of the galactic system discussed by Ludendorff<sup>1</sup> we find a mean amplitude of 1.23. For 232 variables in Theile's compilation<sup>2</sup> we compute a mean photographic amplitude of 1.22.

2. The study of the Cepheid variable stars in the Small Magellanic Cloud has provided material that is uncommonly homogeneous with respect to magnitude system, and therefore suitable for the correlation of amplitude with period and with the median apparent magnitude. All the sequences and all the measures have been reduced to a single sequence and a single observer. Moreover, the number of variables is so large that the correlation can be examined separately for the different parts of the Cloud.

In an earlier paper<sup>3</sup> we have summarized the amplitudes for the classical Cepheids in the Small Cloud, and noted that the background of faint stars in the denser parts of the nucleus diminishes the observable amplitudes on the average by a seventh of a magnitude. Elsewhere in the Cloud the "doubling" effect can be ignored.

3. In table 1, the amplitudes are grouped into means for the 562 Cepheid variables in the Small Cloud—all those for which periods and light curves had been derived from Harvard plates at the time of this compilation. The results are plotted in figure 1. The vertical line through each plotted point has a length double the corresponding mean error of the mean amplitude. The larger amplitudes for the variables with periods in excess of 10 or 12 days have long been known from the earlier evidence from the Magellanic Clouds as well as from the Galactic System. It is verified for galactic Cepheids in Theile's<sup>2</sup> analysis, which is represented also in figure 1 by the

plotted crosses and the broken lines. It seems unlikely, however, that the amplitudes for galactic Cepheids are systematically greater than that for

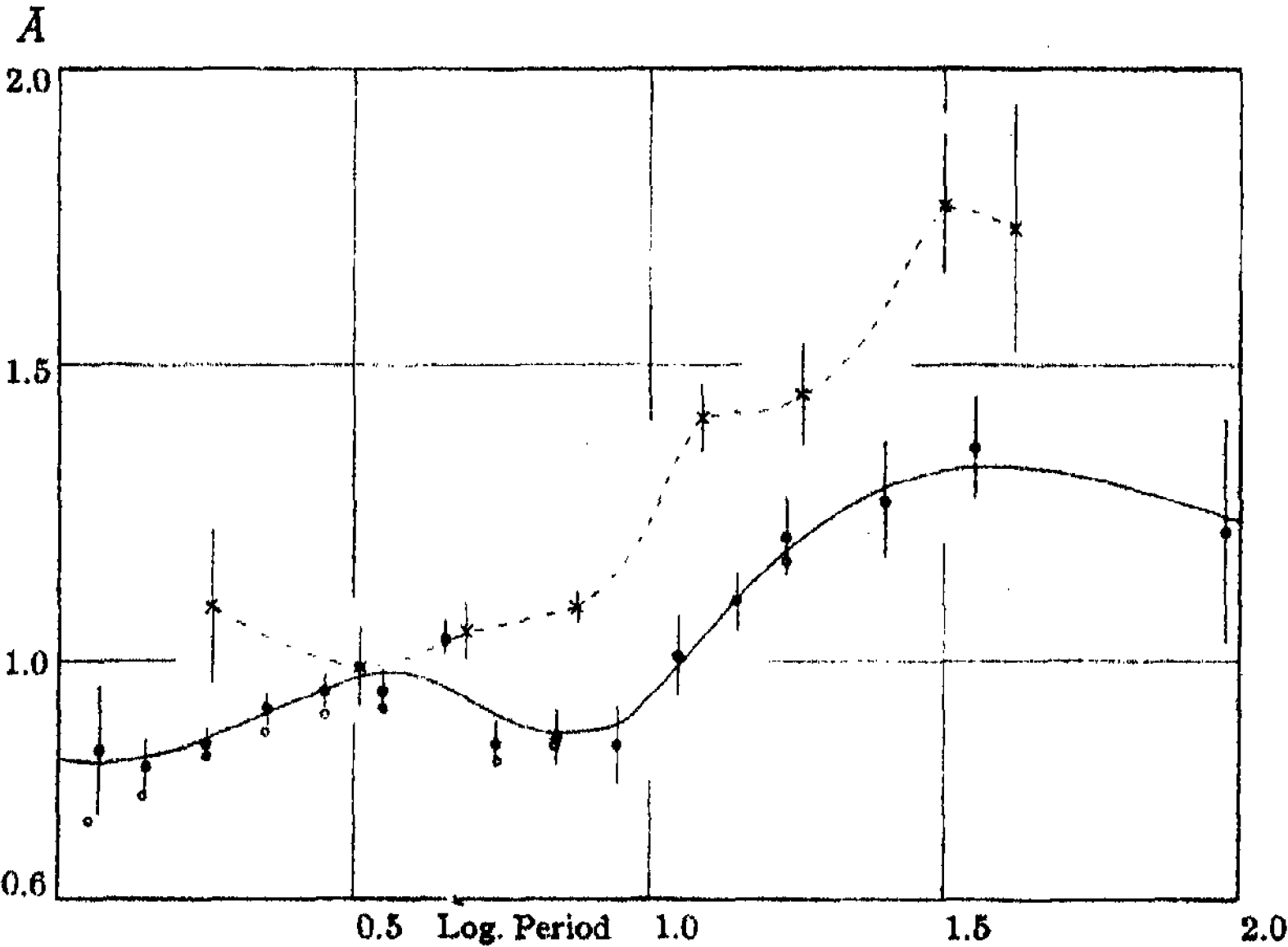


FIGURE 1  
Mean amplitudes and logarithms of periods for Small Magellanic Cloud (dots) and Galaxy (crosses, Theile).

TABLE 1						
PERIOD AND AMPLITUDE						
INTERVAL OF LOG. PERIOD	MEAN LOG. PERIOD		MEAN AMPLITUDE		MEAN ERROR	NUMBER OF VARIABLES
0.0-0.1	0.071	(0.051)	0.85	(0.73)	$\pm 0.11$	5
0.1-0.2	0.148	(0.145)	0.82	(0.77)	0.05	37
0.2-0.3	0.252	(0.251)	0.86	(0.84)	0.03	67
0.3-0.4	0.355	(0.349)	0.92	(0.88)	0.03	58
0.4-0.5	0.454	(0.451)	0.95	(0.91)	0.03	90
0.5-0.6	0.549	(0.549)	0.95	(0.92)	0.04	69
0.6-0.7	0.656	(0.655)	1.04	(1.04)	0.03	53
0.7-0.8	0.740	(0.742)	0.86	(0.83)	0.04	34
0.8-0.9	0.844	(0.842)	0.87	(0.86)	0.05	33
0.9-1.0	0.947	...	0.86	..	0.07	24
1.0-1.1	1.051	...	1.01	..	0.07	21
1.1-1.2	1.149	....	1.10	..	0.05	23
1.2-1.3	1.233	(1.233)	1.21	(1.17)	0.07	16
1.3-1.5	1.401	...	1.27	..	0.10	16
1.5-1.7	1.554	...	1.36	..	0.09	11
>1.7	1.978	...	1.22	..	$\pm 0.19$	6

the Magellanic Cloud by the amount indicated by Theile's results. The matter merits further investigation, not only in the study of the light curves

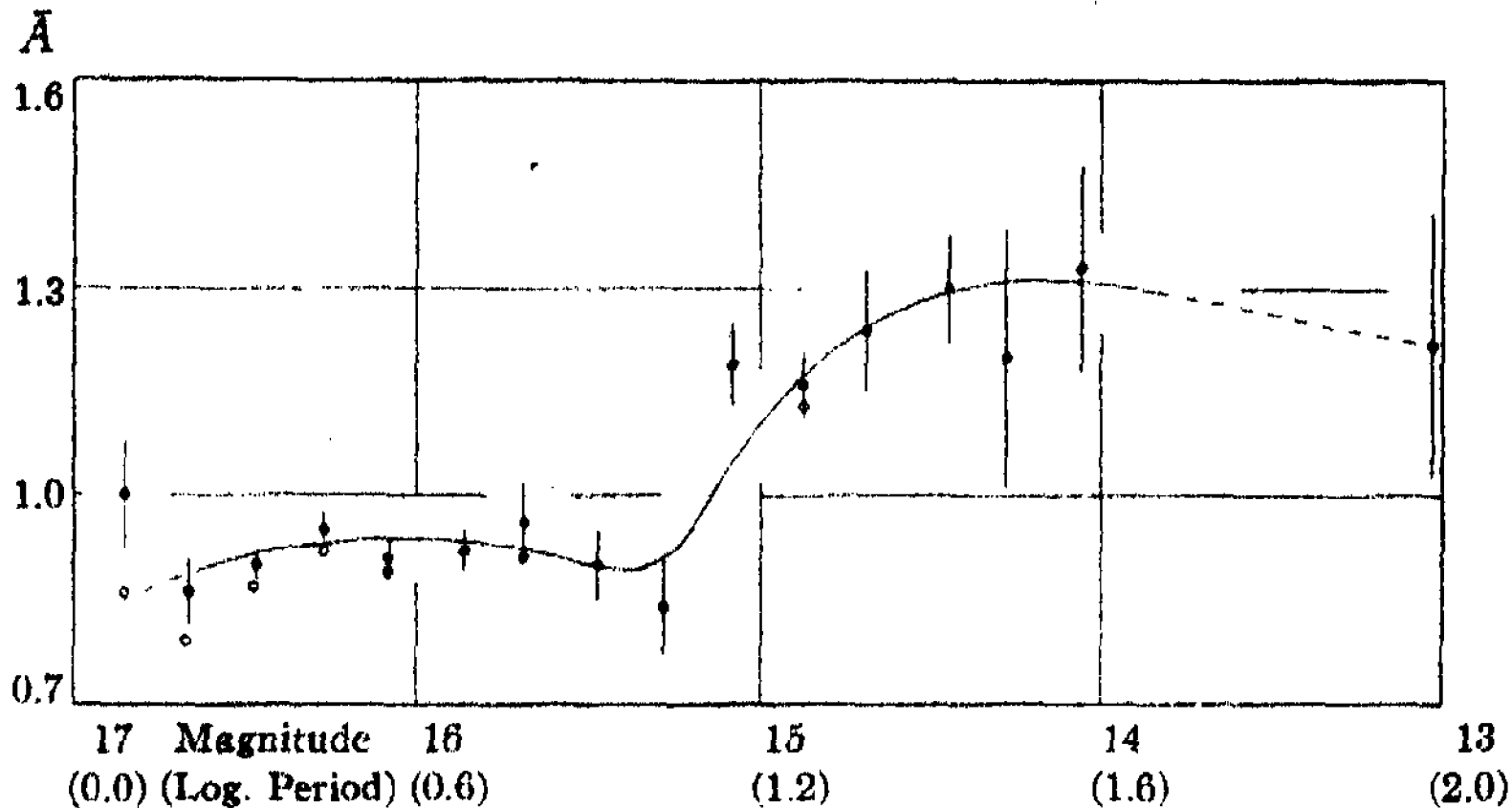


FIGURE 2

Luminosity-amplitude relation for 562 Cepheids in the Small Cloud. Ordinates are mean amplitudes.

TABLE 2  
LUMINOSITY AND AMPLITUDE

INTERVAL OF APP. MAGNITUDE	MEAN MAGNITUDE		MEAN AMPLITUDE		MEAN ERROR	NUMBER OF VARIABLES
	<i>m</i>	<i>m</i>	<i>m</i>	<i>m</i>		
17.00-16.80	16.85	(16.85)	1.00	(0.86)	±0.08	18
16.79-16.60	16.66	(16.87)	0.86	(0.79)	0.05	45
16.59-16.40	16.46	(16.47)	0.90	(0.87)	0.02	92
16.39-16.20	16.27	(16.27)	0.95	(0.92)	0.03	96
16.19-16.00	16.08	(16.08)	0.91	(0.89)	0.03	86
15.99-15.80	15.86	...	0.92	..	0.03	63
15.79-15.60	15.69	(15.69)	0.96	(0.91)	0.06	30
15.59-15.40	15.47	...	0.90	..	0.05	40
15.39-15.20	15.28	...	0.84	..	0.07	17
15.19-15.00	15.08	...	1.19	..	0.06	20
14.99-14.80	14.87	(14.87)	1.16	(1.13)	0.05	16
14.79-14.60	14.69	...	1.24	..	0.09	8
14.59-14.40	14.45	...	1.30	..	0.08	10
14.39-14.20	14.28	...	1.20	..	0.19	7
14.19-14.00	14.06	...	1.33	..	0.15	8
]14.00	13.03	...	1.22	..	±0.19	6

of galactic Cepheids but also through further check of the magnitude system of the Small Cloud.

The reality, for the Small Cloud variables, of the preliminary maximum in the period-amplitude curve for periods between 2.5 and 5 days is not very

certainly established, notwithstanding the indicated small mean errors in the mean amplitudes.

In the determination of the periods and light curves of the 562 Cepheids used in table 1, and in the next section, 85 other variables were measured on all available plates, and preliminary median magnitudes and amplitudes are available, although the periods can as yet only be roughly estimated from the period-luminosity relation. The periods have often been difficult to find because of the small amplitudes. It is therefore appropriate to include these provisional results in a study of the period-amplitude relation and the luminosity-amplitude relation discussed below. The parenthetical entries in the tables, and the open circles in the figures, refer to the revised means that include the 85 "unsolved" Cepheids.

4. The high correlation between the absolute magnitudes and the periods for classical Cepheids (the period-luminosity relation) suggests that a luminosity-amplitude relation should be closely similar to the period-amplitude relation. In table 2 and figure 2 the data for the 562 Cepheids are reassembled to examine this point. The question of the correctness of the magnitude system in the Small Cloud is, of course, also involved in the luminosity-amplitude relation. Assuming its reliability, we find that the amplitudes are, as expected, systematically smaller for the absolutely fainter variables. When the median luminosities exceed absolute magnitude  $-2.2$ , corresponding to apparent magnitudes brighter than 15.15 and periods greater than 13 days, the mean amplitude averages  $1^m.22$ , while for fainter classical Cepheids it is three-tenths of a magnitude less. The brighter Cepheids are redder, but no reliable data are available to test the probable connection between amplitude and color, that is, between amplitude and the wave-length of the maximum in the spectrum-energy curve. The bolometric amplitudes would likely show a somewhat different relation with period and luminosity.

5. Only about one-third of the known variables in the Small Cloud have been measured for amplitude and period. The possibility that the systematic selection of large ranges, or of periods of convenient length, has affected the results summarized in table 1 and figure 1 leads to the necessity of examining the correlation on the basis of the material that is available from those regions of the Cloud that have been thoroughly studied. Table 3 and figure 3 present the results for the 365 variables located in such regions. The relation of mean amplitude to period is very similar to that found for the whole Cloud. Again the suspected Cepheids of unmeasured period are included to form the parenthetical means, which are again plotted as open circles.

Since these data appear to be free of selection, at least for periods greater than two days, we should adopt table 3 and figure 3 as best representing the correlation in the Small Magellanic Cloud.

6. Recognizing the danger of subdividing the observational material too much, we give, nevertheless, in table 4, the mean amplitudes separately for the nuclear, intermediate and border regions of the Cloud, and in each for three intervals of period length. Only the material (365 variables) from the completely surveyed regions is used.

Small amplitudes are found for short period variables in the border regions—possibly an effect on measured magnitudes of distance from plate center. The larger ranges for the longest periods in the same regions may

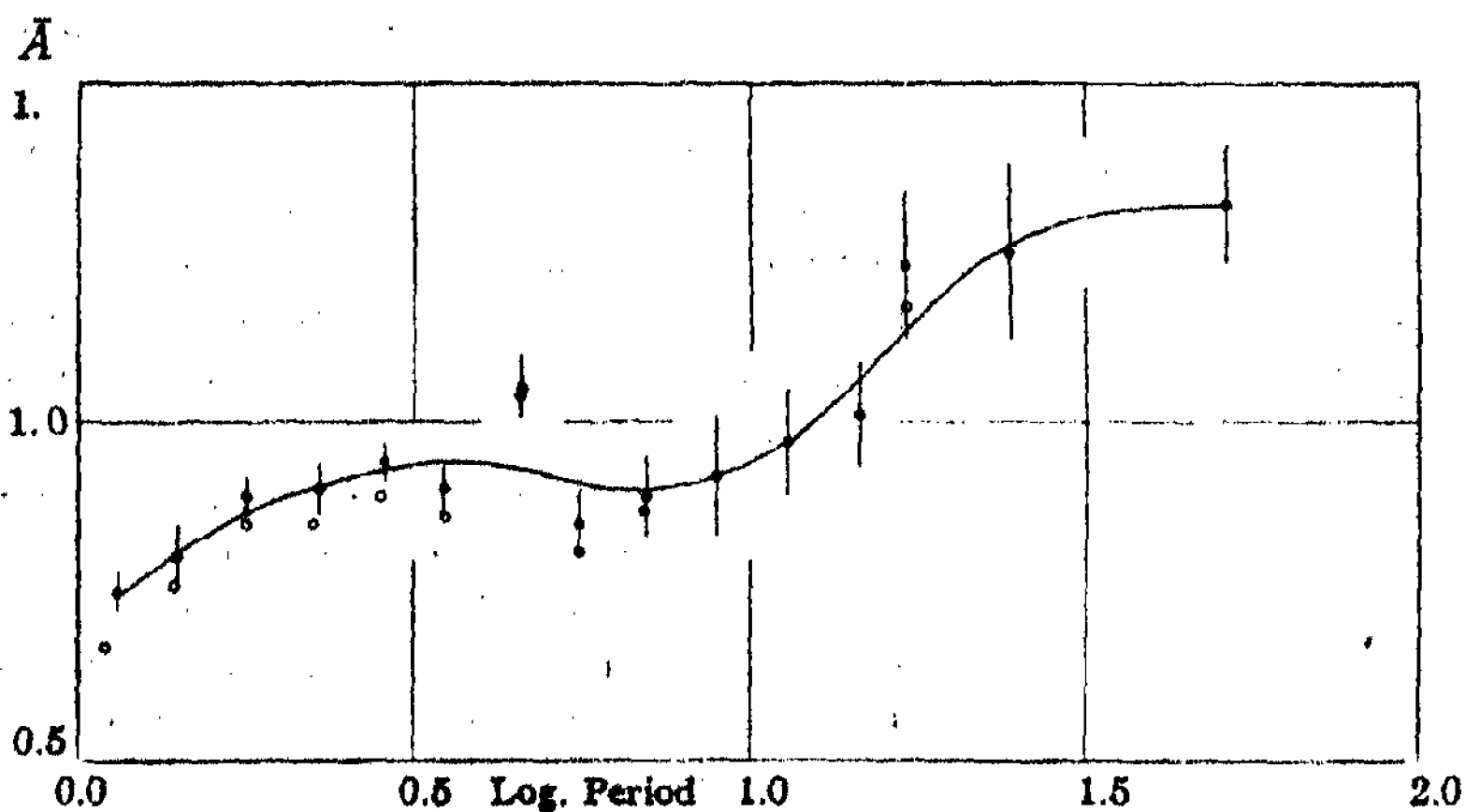


FIGURE 3

Amplitudes and periods for special regions.

TABLE 3  
PERIOD AND AMPLITUDE IN FULLY EXPLORED REGIONS

INTERVAL OF LOG. PERIOD	MEAN LOG. PERIOD		MEAN AMPLITUDE		MEAN ERROR	NUMBER OF VARIABLES
			<i>m</i>	<i>m</i>	<i>m</i>	
0.0-0.1	0.059	(0.043)	0.75	(0.67)	$\pm 0.03$	3
0.1-0.2	0.149	(0.145)	0.80	(0.76)	0.05	29
0.2-0.3	0.252	(0.251)	0.89	(0.85)	0.03	46
0.3-0.4	0.360	(0.351)	0.90	(0.85)	0.04	40
0.4-0.5	0.456	(0.451)	0.94	(0.89)	0.03	63
0.5-0.6	0.545	(0.546)	0.90	(0.86)	0.04	43
0.6-0.7	0.659	(0.657)	1.05	(1.04)	0.05	28
0.7-0.8	0.745	(0.747)	0.85	(0.81)	0.05	21
0.8-0.9	0.845	(0.843)	0.89	(0.87)	0.06	23
0.9-1.0	0.949	...	0.92	..	0.09	18
1.0-1.1	1.055	...	0.97	...	0.08	13
1.1-1.2	1.164	...	1.01	..	0.08	14
1.2-1.3	1.233	(1.233)	1.23	(1.17)	0.11	9
1.3-1.5	1.386	...	1.25	..	0.13	10
>1.5	1.712	...	1.32	...	$\pm 0.09$	5

TABLE 4  
MEAN PERIOD AND AMPLITUDE IN VARIOUS REGIONS<sup>a</sup>

INTERVAL OF LOG. PERIOD	NUCLEUS			INTERMEDIATE			BORDERS			ALL	M. E.
	NO.	$\bar{A}$	M. E.	NO.	$\bar{A}$	M. E.	NO.	$\bar{A}$	M. E.		
<0.4	25	0.91	±0.05	53	0.91	±0.03	40	0.80	±0.03	118	±0.02
0.4-0.8	36	0.94	±0.04	75	0.93	±0.03	44	0.95	±0.05	155	±0.02
>0.8	36	0.95	±0.05	19	1.01	±0.07	37	1.13	±0.06	92	±0.04
All	97	0.93	±0.03	147	0.93	±0.02	121	0.95	±0.03	365	±0.01

<sup>a</sup> This tabulation does not include the 85 suspected Cepheids in the regions.

perhaps owe something to the same effect. The point can be readily checked by using series of plates appropriately centered; such a series will be made at the Boyden Station.

The mean amplitude for all 365 variables is 0<sup>m</sup>.94, a value so low compared with the values derived for galactic Cepheids, 1<sup>m</sup>.22, that, as intimated above, the presence of systematic errors must be suspected. They might be attributed to one or more of the following sources:

(a) Background effect in the Small Cloud (see fourteenth paper of this series). This appears to be excluded by the evidence of the means in the last line of table 4; the average amplitude in the crowded nucleus is not less than elsewhere.

(b) The magnitude system in the Cloud, which has been made homogeneous, and checked for scale as well as zero point. A new observational program is in progress, designed to connect the master magnitude sequence of the Magellanic Clouds with Selected Area 94 at declination 0° and, with the coöperation of Dr. Baade of Mount Wilson, with the International magnitude standards at the North Pole.

(c) Length of photographic exposure. This factor could affect the amplitude of only the shortest periods, since the exposures on the plates used at Harvard for the Magellanic Clouds very rarely exceed one hour.

(d) Incompleteness of light curves for galactic variables, which generally leads to estimating amplitudes from the extreme maximum and minimum observations and not from the more trustworthy mean light curves.

(e) Untrustworthy magnitude sequences, preferential selection of large-range variables, and general non-homogeneity of observations for galactic Cepheids, as mentioned in section 1 above.

The problem is worth further serious consideration because of the persistent question

as to the identity of the laws of Cepheid variation from one galaxy to the next.

Finally it should be noted that for any given period and absolute magnitude there is a large dispersion in the amplitudes of classical Cepheids—a dispersion that far exceeds the observational uncertainties. Ranges of 0.5 and 1.5 magnitudes occur all along the period-amplitude curve. In this paper we have been concerned only with the mean amplitudes, and our conclusions refer only to the general tendency of amplitudes to depend on length of period or on absolute luminosity.

<sup>1</sup> *Sitzber. preuss. Akad. Wiss., Physik-math. Klasse*, 5, 75-78 (1939).

<sup>2</sup> *Veröff. Berlin-Babelsberg*, 12, III, 36 (1939).

<sup>3</sup> These PROCEEDINGS, 28, 195, 197 (1942); the fourteenth paper of this series.





# PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES

Volume 28

December 15, 1942

Number 12

Copyright 1943 by the National Academy of Sciences

## *SEXUAL HORMONES IN ACHLYA. V. HORMONE A', A MALE-SECRETED AUGMENTER OR ACTIVATOR OF HORMONE A*

BY JOHN R. RAPER\*

DEPARTMENT OF BOTANY AND BACTERIOLOGY, INDIANA UNIVERSITY

Communicated November 2, 1942

The rôle of specific sexual hormones in the sexual reaction of two heterothallic species of *Achlya* has been described in earlier papers of this series.<sup>1</sup> Four specific substances were postulated, two secreted by the ♀, bringing about specific reactions in the ♂, and two secreted by the ♂, inducing specific responses in the ♀. On the basis of evidence available at that time it seemed that the entire sexual reaction was initiated with the secretion of hormone *A* by the vegetative mycelium of the ♀, which induced the formation of sexual organ initials, antheridial branches, on the ♂. In the course of a more detailed analysis of the response of the ♂ plant to hormone *A* from the ♀,<sup>2</sup> it was found that the vegetative mycelium of the ♂ secreted some factor which profoundly influenced its response to hormone *A* from the ♀. This factor, produced by the vegetative ♂ plant independently of any response to the ♀, will hereafter be referred to as hormone *A'* (*A* prime).

The species and sexual strains used in this work are those previously employed in the work on sexual hormones in *Achlya*: *Achlya bisexualis* ♀ for the production of hormone *A*,<sup>3</sup> and *Achlya ambisexualis* ♂ as test plants for the assay of hormone *A* and for the detection of hormone *A'*. The method of testing is that previously described<sup>2</sup> and the standard hormone *A* solution is the same as that used throughout the work on the characterization of this hormone.<sup>4</sup> Any modifications of previously described methods or techniques will be described in the body of the paper.

The discovery of hormone *A'* came about quite by accident. In the analysis of the response of the ♂ plant to hormone *A*, a series of experiments was performed in an attempt to determine whether the hormone was used up, i.e., removed from the test solution during the course of the reaction, or whether it produced its effect through the catalytic activation of some existing physiological system, the hormone thus remaining un-

changed and unconsumed in the test solution. The first of these experiments was the repeated quantitative determinations of the hormone *A* content of a 10-ml. sample of standard (6 units/ml.) test solution at successive times with new ♂ plants. The results were unexpected, for instead of the response remaining constant or decreasing there was an *apparent* increase in the amount of hormone *A* during the first day of the experiment. This increased activity reached approximately 300% of that of the sample at the beginning of the experiment and of that of a control series of new samples of hormone *A* solution tested at the same intervals. Toward the end of the second day the reaction decreased somewhat, probably because of the dilution of the hormone *A* solution by water unavoidably carried in with each new pair of ♂ plants. The results of these tests could not be explained at the time.

A second series of experiments to determine the ultimate fate of hormone *A* was then performed. Into 10-ml. samples of standard hormone *A* solution in Petri-plates were placed ♂ test plants with two plants in the first plate, four in the second and so on up to twenty plants in the tenth plate. At the end of two hours the average number of antheridial branches per vegetative hypha was determined for the plants in each sample. On three successive days, October 7, 8 and 9, 1940, the results were the same. In each series the intensity of the response varied directly with the number of test plants in the 10-ml. sample of standard hormone *A* solution. The average responses for the three series are given in table 1.

TABLE 1

MASS EFFECT OF ♂ PLANTS ON THEIR RESPONSE TO HORMONE *A*

Number of plants in tests	2	4	6	8	10	12	14	16	18	20
Average number of antheridial branches	5.2	5.8	9.7	9.8	10.6	13.1	13.7	12.5	13.9	14.1

The results from these tests agree strikingly with those of the previous experiment when, in effect, additional plants were added to the sample.

From these experiments it was obvious that the ♂ plants themselves exerted some mass effect on their response to hormone *A*, and that the intensity of their response depended on some factor in addition to that of hormone *A* concentration. The most probable explanation of this increase seemed to be the secretion by the ♂ plant of a factor which augmented the effect of hormone *A*, or which behaved as an activator for hormone *A*. The secretion of such an active substance, hormone *A'*, has been repeatedly confirmed by experiments extending over a period of two years.

The first tests designed to demonstrate the secretion and activity of an active substance from the ♂ were made as follows. The liquid from a number of 48-hour cultures, each containing ten ♂ plants in 10 ml. of water, was filtered, and this filtrate was used in making up a series of test

solutions containing undiluted  $\sigma^7$  filtrate plus concentrations of hormone *A* over the range of 0-6 units/ml. A second series, similar to the first but made with distilled water instead of  $\sigma^7$  filtrate, served as controls. Into 10-ml. samples of each of these solutions were placed two  $\sigma^7$  test plants which were washed in several changes of distilled water immediately before use. At the end of two hours counts of the antheridial branches were made. The results are given in table 2. A number of facts are apparent when these data are analyzed: (1) The filtrate from the vegetative  $\sigma^7$  plants contains an active substance, hormone *A'*. This sub-

TABLE 2

EFFECT OF  $\sigma^7$  FILTRATE ON THE REACTION OF  $\sigma^7$  PLANTS TO HORMONE *A*. REACTION INTENSITY IS EXPRESSED AS AVERAGE NUMBER OF ANTHERIDIAL BRANCHES

Hormone <i>A</i> , units/ml.	0	1	2	3	4	5	6
Control (no $\sigma^7$ filtrate)	0	0.6	1.7	2.7	3.0	3.6	4.2
$\sigma^7$ filtrate added	0	11.8	12.6	13.1	12.4	10.4	12.5

stance (*a*) in the absence of hormone *A* is inactive in inducing the formation of antheridial branches, but (*b*) in the presence of hormone *A* activates this hormone or augments its activity. (2) The  $\sigma^7$  filtrate, in the presence of hormone *A* within the range of concentrations here used, determines the intensity of the  $\sigma^7$  reaction.

Comparison of the results in the foregoing experiments with those to follow will show that the amount of activation of  $\sigma^7$  filtrate varies appreciably among different samples of  $\sigma^7$  filtrate. Preliminary attempts to standardize the conditions of hormone *A'* production and to establish a quantitative test for its relative concentration have been unsuccessful. The experiments to follow, however, will show something of the activity of hormone *A'* and its relation to the rôle of hormone *A* in inducing the formation of antheridial branches on the  $\sigma^7$  plant.

The interaction of hormones *A* and *A'* is best shown in the following experiment. Two series of test solutions were prepared, and all the samples were tested simultaneously with carefully washed  $\sigma^7$  plants. In the first series hormone *A* in concentrations of 0.0, 0.006, 0.06, 0.6, and 6.0 units/ml. were added to 10-ml. samples of undiluted  $\sigma^7$  filtrate as prepared in the preceding experiment. The second series consisted of 10-ml. samples each containing 6.0 units/ml. hormone *A* plus 100, 50, 10 and 0% of  $\sigma^7$  filtrate. Average antheridial branch production (reaction intensity) for each solution was determined at the end of two hours. The results in figure 1 and table 3 clearly show that the number of antheridial branches produced by the  $\sigma^7$  plant varies in direct proportion to the concentration of both hormones *A* and *A'*, and that for a greater than minimal reaction both hormones must be present in adequate quantity. Thus a deficiency in either results in a low intensity reaction. The absolute necessity of

hormone *A* (from the ♀) for the production of antheridial branches on the ♂ plant is again demonstrated by these results.

TABLE 3

INTERACTION OF HORMONE *A* AND HORMONE *A'* IN THE PRODUCTION OF ANTHERIDIAL BRANCHES

Per cent conc. ♂ filtrate	←-----100-----→					50	10	0
Conc. hormone <i>A</i> , units/ ml.	0.0	0.006	0.06	0.6	←-----6.0-----→			
Reaction intensity	0	2.9	6.7	11.0	27.5	20.6	11.5	7.5

The fact that the ♂ plant in the solution in which no ♂ filtrate was added still gave a reaction cannot be interpreted as meaning that the factor secreted by the ♂ plant is not equally essential. It will be recalled from the first two experiments reported above that the ♂ plants secrete sufficient quantity of hormone *A'* *during the two-hour testing period* to affect significantly the intensity of their reaction.

A marked rhythmic variation in the response of ♂ plants to a standard solution of hormone *A* has been described in an earlier paper.<sup>2</sup> No means have been found to change or eliminate this variation through rigid control of the conditions under which the plants are grown and tested. The substance secreted by the ♂ plant was naturally suspected of being responsible for this variation. Since hormone *A'* increases the activity of hormone *A*, is secreted by the ♂ plant and can limit the reaction within a critical concentration range, it seemed entirely possible that such variation resulted from the elaboration and secretion into the test fluid of varying amounts of the hormone at different times.

The results of a preliminary experiment seemed to justify this belief. At four-hour intervals carefully washed ♂ plants were placed in 10-ml.

TABLE 4

RELATION OF HORMONE *A'* TO THE RHYTHMIC VARIATION OF THE RESPONSE OF ♂ PLANT TO HORMONE *A*, EXPRESSED AS AVERAGE NUMBER OF ANTHERIDIAL BRANCHES

	TIME (NOV. 18-19, 1940)				
	10 A. M.	2 P. M.	6 P. M.	10 P. M.	2 A. M.
Standard hormone <i>A</i> , 6 units/ml.	24.7	5.9	12.5	14.1	16.1
Standard hormone <i>A</i> + ♂ filtrate	24.9	21.7	..	21.8	21.7

samples of (1) standard hormone *A* solution (6 units/ml.) and (2) ♂ filtrate to which had been added hormone *A* in equal concentration. Counts of antheridial branches were made at the end of the four-hour intervals. These tests were performed under the conditions earlier described for the biological assaying of hormone *A*.<sup>2, 4</sup> The average number of antheridial branches produced in each of the solutions at four-hour intervals is given in table 4. In the series using only hormone *A* the coefficient of variation

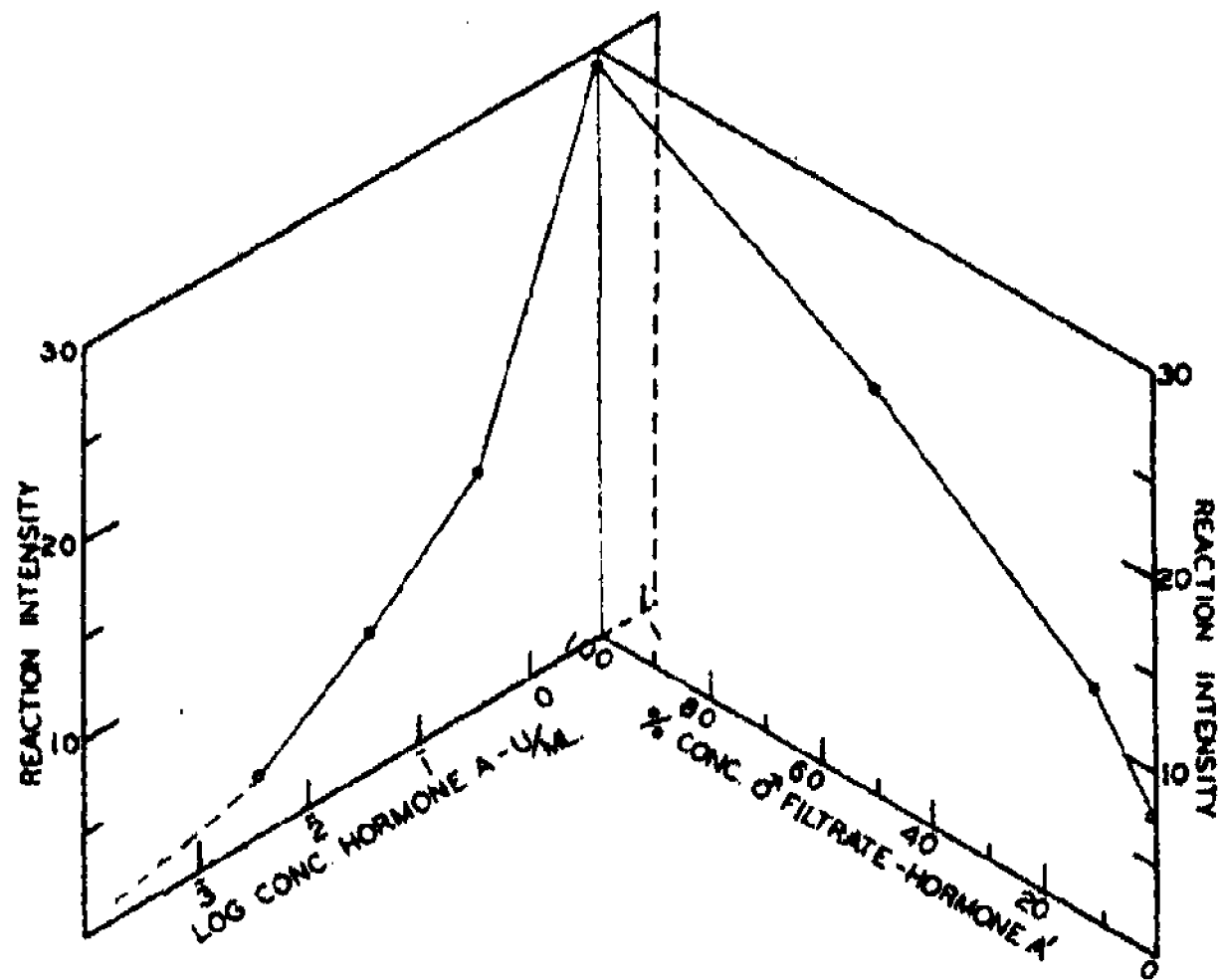


FIGURE 1

Interaction of hormones *A* and *A'* in the production of antheridial branches.

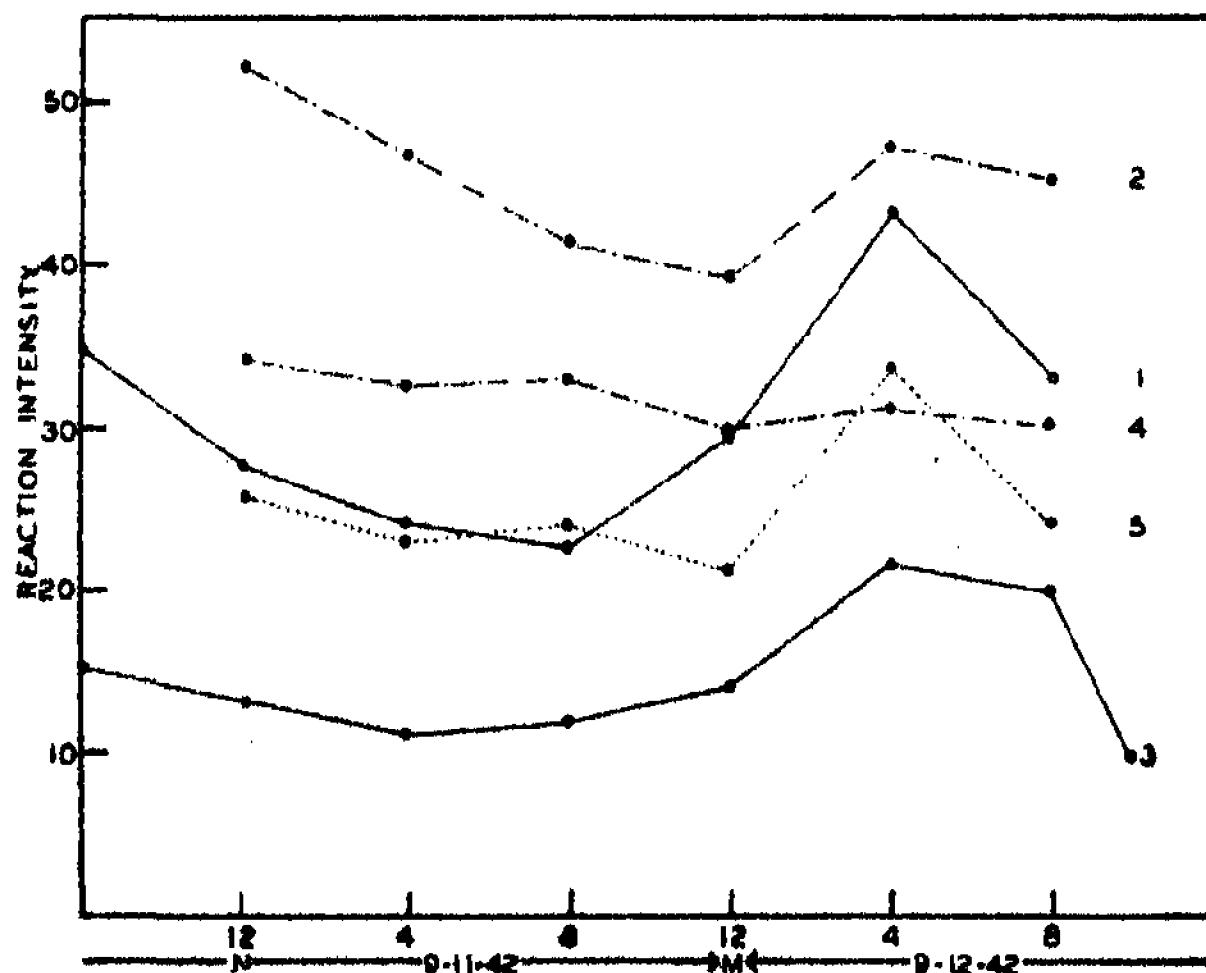


FIGURE 2

Effect of male filtrate on the variation in response of male plants to hormone *A*. The curves represent the reaction intensities of male plants in the following solutions: (1) 60 units of hormone *A*/ml. in distilled water, (2) 60 units hormone *A*/ml. in ♂ filtrate, (3) 6 units hormone *A*/ml. in distilled water, (4) 6 units hormone *A*/ml. in ♂ filtrate, and (5) 6 units hormone *A*/ml. in samples of ♂ filtrate collected at the times indicated.

in the reaction of the ♂ plants was  $50.5 \pm 4.25\%$ . In the series containing added ♂ filtrate the coefficient of variation was reduced to  $25.1 \pm 2.1\%$ .<sup>5</sup>

Later a more extensive and more adequately controlled experiment was conducted to retest the results obtained in the preliminary experiment above. Four solutions were made up as follows: (1) 60 units of hormone *A*/ml. in distilled water, (2) 60 units of hormone *A*/ml. in undiluted filtrate from ♂ plants as described above, (3) 6 units hormone *A*/ml. in distilled water and (4) 6 units of hormone *A*/ml. in ♂ filtrate. Two carefully washed ♂ plants were placed in a 10-ml. sample of each of these solutions, and at the end of four hours counts were made of the antheridial branches on the ♂ plants in each solution. This procedure was repeated at four-hour intervals during 24 hours. In addition to the four series thus established, a fifth series (5) was set up as follows: At the beginning of the experiment twenty ♂ plants were carefully washed and placed in 10 ml. of distilled water. At the end of each four-hour period the liquid was poured off these plants and replaced by fresh distilled water. The samples which were poured off were saved, and at the end of the 24-hour testing period hormone *A* was added to each in a concentration of 6 units/ml. The six samples, comprising series 5, were then tested with ♂ plants simultaneously with a standard (6 units/ml.) hormone *A* solution as control. The results of these five series are plotted in figure 2. It is obvious on examination of these data that under these conditions the variation of ♂ response is not eliminated by the addition of hormone *A*', but in each of the series containing added ♂ filtrate the variation is considerably less than in the corresponding control series lacking ♂ filtrate. The mean and the coefficient of variation in each of the different series are presented in table 5. In series 2, containing ♂ filtrate, the amplitude of the variation is only half that of series 1, which contains the same concentration of hormone *A*. The amplitude of the variation in series 4 is approximately one-fourth that in the corresponding control series lacking added ♂ filtrate. These results taken with those in table 3 demonstrate that the variation of the ♂ response is due at least in part to the quantity of some active factor

TABLE 5  
HORMONE *A*' CONTROL OF RHYTHMIC VARIATION IN RESPONSE OF ♂ PLANTS TO  
HORMONE *A*

SERIES	MEAN	COEFF. OF VARIATION—%
1. 60 units/ml. <i>A</i>	$30.2 \pm 0.81$	$31.6 \pm 2.08$
2. 60 units/ml. <i>A</i> + ♂ filtrate	$45.6 \pm 0.76$	$18.2 \pm 1.21$
3. 6 units/ml. <i>A</i>	$14.9 \pm 0.52$	$42.6 \pm 3.42$
4. 6 units/ml. <i>A</i> + ♂ filtrate	$31.9 \pm 0.47$	$16.5 \pm 1.11$
5. 6 units/ml. <i>A</i> + filtrate from 20 ♂ plants at 4-hour intervals	$25.4 \pm 0.72$	$30.5 \pm 2.28$

secreted by the ♂ mycelium. The rôle of the ♂ secretion is even more strikingly shown in the results of series 5. The curve of the intensity of the reactions induced in the four-hour samples of ♂ filtrate follows rather closely the curves of series 1 and 3, to which no ♂ filtrate was added. The agreement in shape among these curves demonstrates conclusively that the ♂ plant secretes at different times varying quantities of hormone *A'*, and that this variation in amount of secretion results in the rhythmic variation in the response of the plants to a constant concentration of hormone *A*.

Nothing is known as yet about the nature of the mechanism of hormone *A'* effect. Elucidation of the nature of this effect will be difficult, since the only means of detecting the presence and activity of the substance lies in the reaction of the plant which secretes it. Nor is it possible at the present time to say whether or not the substance is indispensable for the formation of antheridial hyphae. It seems entirely possible that it is essential, since carefully washed ♂ plants give a reaction of very low intensity at the time of low hormone *A'* productivity. Before these problems can be solved some means must be found either to inhibit the production of *A'* or to block its effect. Working out a method for the quantitative determination of the active substance will be similarly handicapped. On the other hand, it is difficult to explain the tremendous variation in response of ♂ plants to hormone *A* on the basis of quantitative differences of a single substance, and it is entirely possible that more than a single active factor is secreted into the medium by the ♂ plant.

*Summary.*—In addition to the four specific substances earlier described as initiating and coördinating the sexual reaction in two heterothallic species of *Achlya*, a fifth active factor, hormone *A'*, secreted by the vegetative ♂ mycelium, has now been demonstrated. The activity of this substance has been shown in the following ways: (1) the response of the ♂ plant to hormone *A* increases (*a*) with successively introduced ♂ plants into a single sample of hormone *A* solution, (*b*) with the number of ♂ plants per unit volume of hormone *A* solution, (*c*) in the presence of ♂ filtrate and (*d*) with increasing concentration of ♂ filtrate; and (2) the variation in the response of ♂ plants to hormone *A* depends upon the rhythmic variation in the quantity of hormone *A'* produced by the ♂ plant.

\* The early portions of this work were conducted in the Kerckhoff Biological Laboratories of the California Institute of Technology, Pasadena, Calif., during the tenure of a National Research Fellowship in Botany.

Report of work carried out with the aid of Works Progress Administration, official project No. 165-1-07-172, Work Project No. N-12165. The author wishes to express his indebtedness to Mr. Harvey W. Tomlin for his invaluable assistance.

<sup>1</sup> Raper, John R., *Science*, 89, 321 (1939); *Amer. Jour. Bot.*, 26, 639 (1939); *Ibid.*, 27, 162 (1940).



<sup>2</sup> Raper, John R., *Amer. Jour. Bot.*, **29**, 159 (1942).

<sup>3</sup> Hormone *A* from *Achlya bisexualis* ♀ has been used in these studies for convenience and in order to obtain as nearly reproducible solutions as possible. Tests have been made with hormone *A* from *Achlya ambisexualis* ♀ with entirely comparable results.

<sup>4</sup> Raper, John R., and Haagen-Smit, A. J., *Jour. Biol. Chem.*, **143**, 311 (1942).

<sup>5</sup> In this and the following experiment the coefficient of variation has been calculated from the total number of individual counts rather than from the averages given in table 4 and figure 2. Each of these values represents the average number of antheridial branches on twenty hyphae chosen at random.

---

## OXIDIZED COTTON, AN IMMUNOLOGICALLY SPECIFIC POLYSACCHARIDE

BY MICHAEL HEIDELBERGER AND GLADYS L. HOBBY

PRESBYTERIAN HOSPITAL, COLUMBIA UNIVERSITY

Communicated October 19, 1942

The immunologically specific polysaccharide of Type III pneumococcus, to which the type-specificity and virulence of this pathogenic microorganism are due,<sup>1</sup> is made up of a chain of aldobionic acid,<sup>2</sup> more particularly, cellobiuronic acid,<sup>3</sup> units. The corresponding capsular carbohydrate of the Type VIII pneumococcus contains the same aldobionic acid unit,<sup>4</sup> and in addition, roughly two glucose molecules for every such unit.<sup>5</sup> The close chemical relationship of these polysaccharides is reflected in their serological behavior: each polysaccharide not only precipitates antisera to the homologous pneumococcus type, but the Type III polysaccharide also precipitates a portion of the antibodies in Type VIII antiserum and the Type VIII polysaccharide throws down a portion of the antibodies in Type III antiserum.<sup>6</sup> Evidence has been given that this cross-reactivity is due to some multiple of the repeating unit containing cellobiuronic acid in common.<sup>5</sup>

It has recently been shown that cotton may be oxidized by means of nitrogen dioxide to products containing varying percentages of carboxyl.<sup>7</sup> These products would therefore contain cellobiuronic acid units, separated by glucose, at intervals in the long cellulose chain and might not unreasonably be expected to react specifically with Type VIII antipneumococcus horse serum, or even with the Type III antiserum. Samples of the oxidized cotton containing 16 and 21 per cent of  $-\text{CH}_2\text{OH}$  oxidized to  $-\text{COOH}$  were kindly supplied by Drs. Yackel, Unruh and Kenyon of the Eastman Kodak Laboratories. While it was possible to fractionate the material to some extent, it was most convenient to dissolve weighed quantities in excess  $N$   $\text{NaHCO}_3$  solution, a slow process in the case of the sample of lower  $-\text{COOH}$  content, neutralize the excess of bicarbonate and dilute, first with

water to approximately 0.15 *N* Na<sup>+</sup> content and then with 0.15 *N*, or physiological, NaCl solution. Qualitative tests showed that both samples, at high dilutions, precipitated antipneumococcus Types III and VIII horse sera, and failed to react with Type I and Type II antisera. Quantitative data<sup>8</sup> for solutions of the sample with 16% —COOH indicated a degree of cross-reactivity entirely comparable to that of the corresponding pneumococcus specific polysaccharides.<sup>5, 9</sup> Moreover, the antibody fraction reactive with the oxidized cotton was mainly the portion cross-reactive with the pneumococcus polysaccharide of heterologous type.

Analyses, calculated to 1.0 ml. portions, with Type VIII antipneumococcus horse serum 909, containing 1.0 mg. of Type VIII anticarbohydrate nitrogen per ml.:<sup>10</sup>

POLYSACCHARIDE ADDED	QUANTITY OF POLYSACCHARIDE, MG.	ANTIBODY NITROGEN PRECIPITATED, MG.
Type III pneumococcus	0.13	0.202
Type III pneumococcus	0.4	0.278
Oxidized cotton (16% —COOH)	0.02	0.158
Oxidized cotton (16% —COOH)	0.07	0.273

The supernatant from the last tube, plus 0.2 mg. Type III polysaccharide yielded 0.075 mg. of cross-reactive antibody N, showing that roughly two-thirds of the cross-reactive antibody had been removed by the cotton solution.

Analyses, calculated to 1.0-ml. portions, with Type III antipneumococcus horse serum 792, containing 0.7 mg. of Type III anticarbohydrate nitrogen per ml.:<sup>10</sup>

POLYSACCHARIDE ADDED	QUANTITY OF POLY- SACCHARIDE, MG.	ANTIBODY NITROGEN PRECIPITATED, MG.
Type VIII pneumococcus	0.02	0.111
Type VIII pneumococcus	0.2	0.190
Oxidized cotton	0.008	0.080
Oxidized cotton	0.05	0.165
Oxidized cotton	0.2	0.150

The supernatant from the last tube, plus 0.1 mg. Type VIII specific polysaccharide, yielded only 0.011 mg. of antibody N, showing that most of the cross-reactive antibody had been removed by the cotton solution. The supernatant from the next-to-last tube gave 0.016 mg. antibody N with Type VIII polysaccharide, while that from the 0.008-cotton tube gave 0.065 mg. The sums of the two sets of values are nearly equal to the antibody N precipitated by the Type VIII polysaccharide alone.

As would be expected from its chemical structure, the oxidized cotton with 16% —COOH corresponds more closely in its immunological behavior to the Type VIII substance than to the Type III polysaccharide.

Attempts were also made to confer active protection upon mice with doses of oxidized cotton ranging from 0.005 mg. down to 0.00001 mg.,<sup>11</sup> but the material failed to protect against 10 to 100 minimal lethal doses of an extremely virulent strain of Type VIII pneumococcus. However, the same amounts of the Type III and Type VIII specific polysaccharides also failed to give protection against this strain.

The precipitation data with the oxidized cotton again emphasize the strict correlation between chemical constitution and immunological specificity and show that predictions as to reactivity may be made when the constitution of the repeating unit responsible for that reactivity is known.

<sup>1</sup> Heidelberger, M., and Avery, O. T., *Jour. Exp. Med.*, 38, 73 (1923); 40, 301 (1924). Avery, O. T., and Heidelberger, M., *Ibid.*, 42, 367 (1925).

<sup>2</sup> Heidelberger, M., and Goebel, W. F., *Jour. Biol. Chem.*, 70, 613 (1926); 74, 613 (1927).

<sup>3</sup> Hotchkiss, R. D., and Goebel, W. F., *Ibid.*, 121, 195 (1937). Reeves, R. E., and Goebel, W. F., *Ibid.*, 139, 511 (1941). Adams, M. H., Reeves, R. E., and Goebel, W. F., *Ibid.*, 140, 181 (1941).

<sup>4</sup> Goebel, W. F., *Ibid.*, 110, 391 (1935).

<sup>5</sup> Heidelberger, M., Kabat, E. A., and Shrivastava, D. L., *Jour. Exp. Med.*, 65, 487 (1937).

<sup>6</sup> Sugg, J. Y., Gaspari, E. L., Fleming, W. L., and Neill, J. M., *Ibid.*, 47, 917 (1928). Cooper, G., Edwards, M., and Rosenstein, C., *Ibid.*, 49, 481 (1929). Also Ref. 5. Anti-pneumococcus Types III and VIII horse sera show this cross-precipitation more regularly than do the corresponding rabbit sera.

<sup>7</sup> Yackel, E. C., and Kenyon, W. O., *Jour. Am. Chem. Soc.*, 64, 121 (1942). Unruh, C. C., and Kenyon, W. O., *Ibid.*, 64, 127 (1942).

<sup>8</sup> Heidelberger, M., Kendall, F. E., and Soo Hoo, C. M., *Jour. Exp. Med.*, 58, 137 (1933). Heidelberger, M., and Kendall, F. E., *Ibid.*, 61, 559; 62, 697 (1935).

<sup>9</sup> Heidelberger, M., Kabat, E. A., and Mayer, M., *Ibid.*, 75, 35 (1942).

<sup>10</sup> After removal of antibodies precipitated by pneumococcus C-substance.

<sup>11</sup> Schiemaun, O., and Casper, W., *Zeitschr. f. Hyg.*, 108, 220 (1927). Felton, L. D., *Jour. Infect. Dis.*, 56, 101 (1935).

## THE GENETIC NATURE OF X-RAY INDUCED CHANGES IN POLLEN\*

By CHARLES M. RICK

DIVISION OF TRUCK CROPS, UNIVERSITY OF CALIFORNIA

Communicated October 28, 1942

Ample evidence exists in genetic literature to suggest that many more pollen characters are genetically self-determined in the pollen than have already been reported. Some of these cases—for example, the genes determining small pollen in *Zea*—were discovered by virtue of the fact that

small pollen is less viable and hence causes distorted ratios of more conspicuous characters to which the genes for small pollen size are linked.<sup>1</sup> Some other respects in which genetic autonomy occurs in pollen include: chemical composition,<sup>2</sup> viability in the many instances of semi-sterility and behavior as in the many examples of oppositional self-sterility allelomorphs.

With the known mutation inducing agents available, it should be possible to demonstrate whether or not certain characters are determined in this manner. Since size is usually governed by a large number of genetic factors, each of relatively small effect and each integrated with others in its action, mutation in a single gene of this type in pollen would be very difficult to detect. Mutation of a large number, however, in pollen where the haploid condition would permit expression of recessive as well as dominant changes, should be readily detectable as an increase in variation of size. Accordingly, experiments were performed to detect whether or not x-rays would induce any genetic response in size of pollen. Lethal effects, i.e., changes resulting in abortion of pollen, were also studied.

Genetically self-determined pollen abortion, whether conditioned by chromosomal deficiencies and duplications or by gene mutation, has been observed frequently. There are also reports in the literature of a similar control of microspore and pollen size.<sup>3</sup> Nevertheless, pollen size may also be determined by the sporophyte producing the pollen instead of by the gametophyte itself.<sup>4</sup>

Typical of pollen measurements in general, the intra-treatment variability of sample values was much greater than would be expected on the basis of pure chance variation. In order to cope with this natural fluctuation in comparing different treatments statistically, many samples of small size were taken and  $n$  was taken as the number of samples, instead of the total number of grains observed, as the basis for comparisons between treatments. Pollen abortion, mean length of grain and variance of those lengths were calculated for each sample. Means and standard errors were then calculated for the distribution of these sample values. In the following report a random sample of 25 was adopted for measurement of lengths, and 100 for frequency of pollen abortion.

The pollen collections were examined in aceto-carmin in order to estimate pollen abortion in the same collection used for measurement of lengths. Examination was made immediately after mounting. A pollen grain was considered aborted if it was devoid of cytoplasmic contents as indicated by the aceto-carmin stain. Length of grain is the most satisfactory dimension for measurement of the elongate grains of *Tradescantia* and *Pisum* lines used here. Lengths were measured by means of an ocular micrometer.

*Diploid-Tetraploid Comparisons.*—A comparison between diploid and closely related autotetraploid lines should offer a test of the hypothesis of

x-ray induced mutations in pollen grains. Stadler<sup>5</sup> has demonstrated conclusively that in the cereals the rate of x-ray induced mutation drops rapidly with increasing degree of polyploidy. The pollen produced by autotetraploids is diploid, consisting of two identical sets of chromosomes. Since every gene exists in duplicate, a mutation from the dominant to recessive condition of any gene will gain no phenotypic expression because of the presence of the dominant allele except in the very rare event that the same recessive mutation is induced in both alleles. On the other hand, the haploid pollen produced by diploids should show phenotypic expression of recessive mutations. Since the great majority of x-ray induced mutations are recessive, this comparison should be illuminating.

In the x-ray treatments clones of *Tradescantia* species were used because of the data available on the timing of the cycle of development of the male gametophyte.<sup>6</sup> Collections were made from the eight to eleventh day after the material had been irradiated. In terms of the developmental cycle these collections were taken from buds in which the pollen mother cells had just completed meiosis and the microspores were experiencing the first half of the post-meiotic resting stage at the time of treatment. The conditions of radiation are given in table 1.

TABLE 1

COMPARISON OF X-RAY EFFECTS ON ABORTION OF POLLEN, MEAN AND VARIANCE OF LENGTH OF GRAIN IN DIPLOID AND AUTOTETRAPLOID *Tradescantia* SPECIES

200 kv., 15 ma., 81 cm.,  $\frac{1}{2}$  mm. copper

SPECIES	TREAT- MENT	NUMBER OF SAMPLES	POLLEN ABOR- TION, %	LENGTH OF POLLEN GRAIN MEAN, $\mu$ VARIANCE, $\mu^2$	
<i>T. paludosa</i>					
Diploid	Control	16	4.99 $\pm$ 0.49	49.11 $\pm$ 0.19	0.742 $\pm$ 0.042
	400 r	18	10.98 $\pm$ 0.83	47.89 $\pm$ 0.24	2.163 $\pm$ 0.158
<i>T. canaliculata</i>					
Autotetra- ploid	Control	20	13.10 $\pm$ 0.90	57.24 $\pm$ 0.27	1.322 $\pm$ 0.062
	400 r	18	13.20 $\pm$ 0.79	56.34 $\pm$ 0.45	1.308 $\pm$ 0.138
Clone No. 374*					
Diploid	Control	10	22.2 $\pm$ 1.2	49.38 $\pm$ 0.26	2.549 $\pm$ 0.037
	200 r	10	28.87 $\pm$ 1.43	47.99 $\pm$ 0.14	3.043 $\pm$ 0.149
<i>T. virginiana</i>					
Autotetra- ploid	Control	19	26.03 $\pm$ 0.92	56.38 $\pm$ 0.33	1.879 $\pm$ 0.115
	400 r	19	26.84 $\pm$ 1.25	59.13 $\pm$ 0.20	1.940 $\pm$ 0.088

\* A segregate from the cross, *T. canaliculata* x *T. humilis*.

Two clones of diploid and two of tetraploid *Tradescantia* species were used in this study. Evidence of the autotetraploid nature of the tetraploid species, *T. virginiana* L. and *T. canaliculata* Raf., has been presented by

Anderson and Sax.<sup>7</sup> Of the diploid clones used, one was identified as *T. paludosa* Anderson and Woodson and the other was a segregate from the cross, *T. canaliculata* Raf. x *T. humilis* Rose. The diploid *T. paludosa* and the tetraploid *T. canaliculata* were irradiated simultaneously and grown under the same conditions. The other clones were treated at different times.

The results are summarized in table 1. In the diploid clones the radiation induced a very significant increase in pollen abortion and in variability of length and a decrease in mean length. These changes showed a slight gradual increase during the four-day period of collections, but in comparison with measurements of collections on preceding days, these values are at a new high level and for purposes of comparison can be safely considered as a unit. This pattern of x-ray effect is typical of many that have been investigated here in diploid *Tradescantia*.

This response to x-ray treatment of the haploid microspores would have occurred if the radiation induced mutations (chromosomal aberrations as well as gene mutations in the strict sense) which were immediately expressed in the size and viability of the microspores. The great preponderance of x-ray induced mutations reported in the literature has been of the negative type; therefore, it is no surprise that the changes encountered here have been mostly, if not entirely, in the direction of smaller grains, and that the percentage of aborted grains showed a significant increase. The changes involved will be described in more detail in a later publication.

The effect on the pollen of autotetraploid clones is strikingly different. The rate of pollen abortion and the variance of pollen grain length remain very little affected by radiation which caused quite significant changes in the diploid, thus bearing out the genetic interpretation. Mean length of grain seemed to be affected, in one instance to a nearly significantly lower level and in the other to a very significantly higher level. No explanation would seem to account for these peculiar trends. Of the three values, mean length in general shows the greatest fluctuations and lengths from the same plant may be subject to substantial increases and decreases over a period of time. Since the changes here are both positive and negative, it is doubtful whether they are related to the x-ray treatment.

In a comparison of the x-ray sensitivity of microspore chromosomes, Sax and Swanson<sup>8</sup> found the rate of aberration in chromosomes of diploid *Tradescantia* microspores to be only half that of the haploid. These results were expressed on a chromosome basis; if expressed as frequencies per cell, the sensitivity of haploid and diploid would be about equal. Yet, even if this might indicate a gene sensitivity reduced to one-half in the diploid microspore, the differences in table 1 would remain relatively unaffected. For instance, the negative changes in the autotetraploid clones

must be increased by factors of 3 to 25 before the differences reach the five per cent level of significance.

TABLE 2  
COMPARATIVE EFFECTS OF X-RAYS APPLIED AT 3°C. AND 33°C. ON MEASUREMENTS OF  
POLLEN GRAINS

134 kv., 10 ma., 50 cm., no filters

Ten samples per treatment

TREATMENT NUMBER	RADI- ATION	TEMPERA- TURE, °C.	POLLEN ABORTION, %	LENGTH OF POLLEN	
				MEAN, $\mu$	STANDARD DE- VIATION, $\mu$
20*	150 r	3	11.5 $\pm$ 0.9	47.11 $\pm$ 0.63	2.758 $\pm$ 0.096
21*	150 r	33	10.2 $\pm$ 0.9	47.67 $\pm$ 0.54	2.492 $\pm$ 0.082
24	None	3	9.5 $\pm$ 0.6	46.57 $\pm$ 0.59	1.836 $\pm$ 0.099
19	None	33	.....	46.75 $\pm$ 0.48	2.027 $\pm$ 0.082

\* Lots 20 and 21 were irradiated simultaneously.

*Effect of Temperature on X-ray Induced Changes.*—Only one trial has been made of the effect of temperature. The pollen test of mutation rate offers advantages of rapidity and ease of manipulation for tests of this sort and might prove useful in future experiments where the contributory effects of dosage, intensity, temperature and other factors are studied. Inflorescences of a diploid clone of *Tradescantia paludosa* were irradiated simultaneously in cardboard containers of warm and cold water. The temperature differences were maintained during the period of treatment and for several hours afterward. The results and other conditions of the experiment are given in table 2. Temperature influenced the values significantly only in the case of variation of length and here the *P* value of the difference is 0.036, but the effect of radiation at low temperature is consistently more intense in each of the three measures. From the standpoint of direction and consistency alone these results are well within the realm of chance occurrences and larger samples will be needed from material treated with heavier doses before it can be said with certainty that low temperature enhances the x-ray effect in all respects.

The increased variability induced at 3° over 33°C. would be rather difficult to account for in terms of a physiological response other than a genetic one. The genetic interpretation is admittedly far from satisfactory, yet the point of interest here is that this response in variability of pollen size agrees with the response found in most other experiments, the genetic nature of which is undisputed. For instance, the frequency of x-ray induced chromosomal aberrations in *Tradescantia*<sup>9</sup> and in *Drosophila*<sup>10</sup> is greater when treatments were applied at lower temperatures. A similar response has been found in lethal mutations in *Drosophila*<sup>11</sup> and in chlorophyll deficient mutants in *Hordeum*<sup>12</sup> but these reports are contradicted



by the finding of no temperature effect by others.<sup>13</sup> The picture conveyed by the literature then is confused but there is agreement to the effect that there is no positive temperature coefficient of x-ray induced mutation and in this respect the response in pollen resembles mutation.

*Inheritance of Pollen Size in Pisum sativum L.*—As mentioned above there is ample evidence in the literature of genes governing size, viability and physiological activity of pollen, which segregate at meiosis and gain expression immediately in the grains to which they are contributed. In these cases the segregating types have been sharply distinguished and only one or a few gene pairs have been concerned in their determination. Although it did not seem unreasonable to suppose that pollen size could be regulated by multiple factors, it seemed highly desirable to have some such example in untreated material to compare with the x-ray treatments.

The naturally self-pollinated legumes offer a source of lines which are homozygous yet do not suffer the usual effects of inbreeding. In the garden pea, *Pisum sativum* L., varieties were found which differed in length of pollen grain. Crosses were made between large and small types and the  $F_1$  hybrids were grown simultaneously with selfed progeny from the plants used as parents. For comparative data, samples were taken from the selfed progenies of the parents rather than from the parent plants themselves in order that the parent lines be grown under the same conditions as the hybrids. The homozygous condition of the parent plants attested by the uniformity of their progeny justified this measure.

TABLE 3

MEASUREMENTS OF POLLEN IN  $F_1$  AND PARENTAL LINES OF GARDEN PEA

NUMBER OF LINE	DESCRIPTION	NUMBER OF SAMPLES	POLLEN ABORTION, %	LENGTH OF POLLEN GRAIN MEAN, $\mu$	VARIANCE, $\mu^2$
42 $P_s$ 8	$F_1$ , 3 $\times$ 13	18	2.68 $\pm$ 0.55	50.68 $\pm$ 0.32	1.267 $\pm$ 0.105
42 $P_s$ 18	$F_1$ , 25 $\times$ 26	9	2.49 $\pm$ 0.72	52.13 $\pm$ 0.58	1.800 $\pm$ 0.271
42 $P_s$ 22	$F_1$ , 11 $\times$ 25	19	3.71 $\pm$ 0.44	51.25 $\pm$ 0.46	1.511 $\pm$ 0.075
42 $P_s$ 28	$F_1$ , 31 $\times$ 25	11	1.94 $\pm$ 0.46	51.05 $\pm$ 0.84	1.270 $\pm$ 0.070
42 $P_s$ 3	$P_1$	19	2.42 $\pm$ 0.35	55.58 $\pm$ 0.21	0.602 $\pm$ 0.038
42 $P_s$ 11	$P_1$	18	2.89 $\pm$ 0.30	54.39 $\pm$ 0.29	0.590 $\pm$ 0.040
42 $P_s$ 18	$P_1$	18	5.91 $\pm$ 0.72	49.67 $\pm$ 0.37	0.545 $\pm$ 0.043
42 $P_s$ 25	$P_1$	25	5.27 $\pm$ 0.44	49.57 $\pm$ 0.32	0.685 $\pm$ 0.045
42 $P_s$ 26	$P_1$	20	8.17 $\pm$ 0.39	54.01 $\pm$ 0.14	0.535 $\pm$ 0.023
42 $P_s$ 31	$P_1$	3	10 $\pm$	53.78 $\pm$ 0.42	0.552 $\pm$ 0.050

Ample data for statistical comparisons are available from four different  $F_1$ 's and their parents. The observations are presented in table 3 and figure 1. In every case pollen of the hybrid was intermediate in size between its two parents; furthermore, the variation of the hybrid pollen grain length significantly exceeded that of each parent in every case. In the case of



least difference, i.e., the  $F_1$ , 42 Ps 18 and its parent, 42 Ps 25, the  $t$  value of the difference is 2.63 with a corresponding  $P$  value of less than 0.01. This is precisely the effect expected if size of pollen were to some extent governed by the genotype of the pollen itself and if a large number of genes, each having a small effect, were involved. A possible source of internal instability might be hybridity for chromosome rearrangement, say inversions or translocations. Sufficient buds of the hybrids were not available to permit a cytological study to detect such hybridity but the four pea hybrids show no increased percentage of pollen abortion which is characteristic of inversion and translocation heterozygotes.

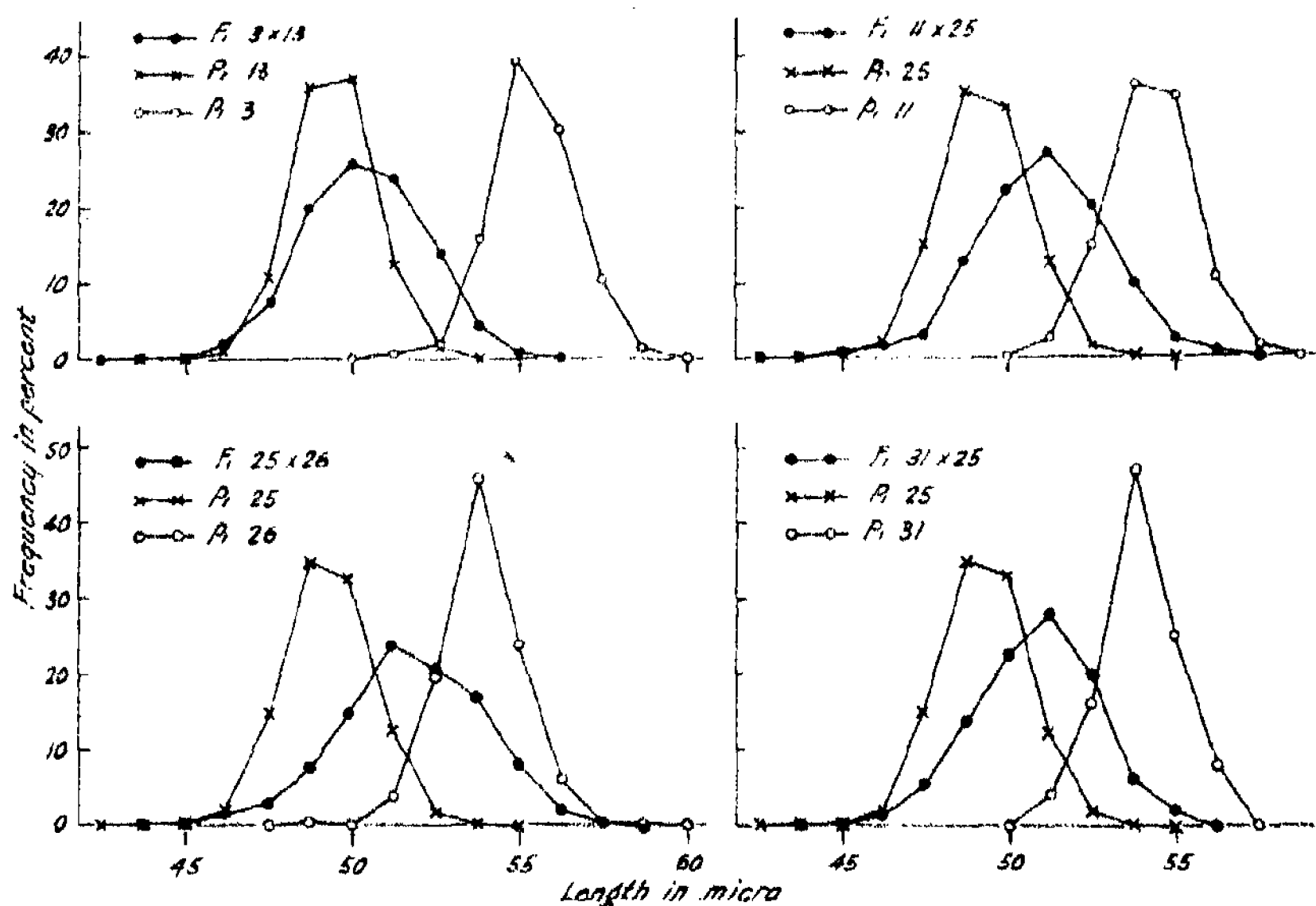


FIGURE 1

Frequency distributions of length of pollen in  $F_1$  and parental lines of garden pea.

Unless a very great number of factors interact to determine size differences in these hybrids, the distribution of lengths in the hybrids might be expected to completely cover the range of either parent. Also, the presence of only a few determiners of small size in the larger parent or vice versa might lead to considerable transgressive variation. Since these  $F_1$  distributions are all contained within the range of either parent and one in particular, 42 Ps 8, fails to cover the entire range of its parents, it is possible that even in these hybrids size is determined maternally to a limited extent. Nevertheless, the consistent tendency of all hybrids to be more variable than either parent certainly points to a multifactorial genetic determination of size within each grain.

*Summary.*—X-rays applied to microspores of diploid species of *Tradescantia* shortly after meiosis cause a significant increase in variability of length of the subsequently developed pollen grains and in the percentage of aborted pollen. There is also a significant decrease in the mean lengths of grains. Similarly treated autotetraploid species of *Tradescantia* show no significant changes. This difference is interpreted to indicate that in the diploid pollen of autotetraploids any recessive mutation is masked by its dominant allele.

X-rays induced significantly greater variability of pollen length when applied to diploid *Tradescantia* at 3° than when applied at 33°C.

In consideration of these results it is concluded that size and viability of pollen are, at least in part, genetically self-determined and that the changes observed are the consequence of mutations induced by the x-ray treatment.

These conclusions are supported by an analysis of genetic variation in pollen length in *Pisum sativum*.

\* The writer takes great pleasure in acknowledging the many constructive suggestions offered by Dr. Karl Sax of Harvard University. Martha O. Rick, the writer's wife, aided to a great extent in many phases of the work, and to her the writer wishes to express his sincerest obligations. The writer is indebted to Dr. P. C. Aebersold of the Crocker Radiation Laboratory for his kind help in applying the x-rays.

<sup>1</sup> Mangelsdorf, P. C., *Jour. Hered.*, **23**, 289-295 (1932); Rhoades, M. M., and Rhoades, Virginia H., *Genetics*, **24**, 302-314 (1939).

<sup>2</sup> Brink, R. A., and MacGillivray, J. H., *Amer. Jour. Bot.*, **11**, 465-469 (1924); Demerec, M., *Ibid.*, **11**, 461-464 (1924); Parnell, F. R., *Jour. Genet.*, **11**, 209-212 (1921).

<sup>3</sup> Mangelsdorf, P. C., loc. cit.; Renner, O., *Zeit. Bot.*, **11**, 305-380 (1919); Satina, Sophia, and Blakeslee, A. F., *Amer. Jour. Bot.*, **24**, 518-527 (1937); Sax, K., *Genetics*, **22**, 523-533 (1937); Sinotô, Y., *Cytologia*, **1**, 109-191 (1929).

<sup>4</sup> Bateson, W., Saunders, E. R., Punnett, R. C., and Hurst, C. C., *Rept. Evol. Com. Roy. Soc. Rept.*, **2**, 154 pp. (1905). Another example of sporophytic control is the difference in size and color of pollen maintained in anthers of different levels in the flower of *Lythrum Salicaria* L.

<sup>5</sup> Stadler, L. J., these PROCEEDINGS, **15**, 876-881 (1929).

<sup>6</sup> Sax, K., *Genetics*, **23**, 494-516 (1938).

<sup>7</sup> Anderson, E., and Sax, K., *Bot. Gaz.*, **97**, 433-476 (1936).

<sup>8</sup> Sax, K., and Swanson, C. P., *Amer. Jour. Bot.*, **28**, 52-59 (1941).

<sup>9</sup> Fabergé, A. C., *Jour. Genet.*, **39**, 229-248 (1940); Rick, C. M., *Genetics*, **25**, 466-482 (1940); Sax, K., and Enzmann, E. V., these PROCEEDINGS, **25**, 397-405 (1939).

<sup>10</sup> Mickey, G. H., *Genetica*, **21**, 386-407 (1939).

<sup>11</sup> Medvedev, N. N., *Compt. Rend. Acad. Sci. U.R.S.S.*, **4**, 283-285 (1935); Medvedev, N. N., *Ibid.*, **19**, 301-302 (1938); Timoféeff-Ressovsky, N. W., *Mutationsforschung in der Vererbungslehre*, 184 pp., Theodor Steinkopf, Dresden (1937).

<sup>12</sup> Rick, C. M. (unpublished).

<sup>13</sup> Stadler, L. J., *Jour. Hered.*, **21**, 3-19 (1930); Stadler, L. J., *Sci. Agr.*, **11**, 645-661 (1931); Timoféeff-Ressovsky, N. W., und Zimmer, K. G., *Biol. Zbl.*, **59**, 358-362 (1939).

## THE LOWER HYDRATES OF SOAP

M. J. BUEGER, L. B. SMITH, A. DE BRETTEVILLE, JR., AND F. V. RYER

Communicated October 24, 1942

1. *Introduction.*—It has been pointed out<sup>1</sup> that the non-recognition of the importance of minor "impurities" may lead to grave misinterpretations of the stability relations between crystalline phases. Conversely, anomalous phase relationships are cause for suspecting that "impurities" may be playing rôles which cannot be neglected. Some of the relationships supposed to exist between certain forms of soap, for example, are almost incredible, and strongly suggest that the soap forms in question do not fall into a simple one-component system, as hitherto supposed, but are parts rather of a two-component system. Our experimental investigation of the relationship between some of these soap forms herein reported indeed proves that the second component is water, and that not a few of the all-too-numerous and supposedly anhydrous soap forms are actually hydrates. This paper is devoted to a discussion of the relation of the lower hydrates and anhydrous soap to one another, together with a brief résumé of the experimental evidence for these relations.

2. *Extension of the Temperature Range of the So-Called  $\alpha$  Form.*—According to Thiessen and Stauff,<sup>2</sup> the form of sodium stearate which precipitates from alcoholic solutions, called by them the  $\alpha$  form, is metastable and, on heating, changes irreversibly to the stable  $\beta$  form at about 52°C. The further relationships between these forms as given by them are most anomalous. These relationships would be understandable if the two forms were not the same compound. This suspicion is strengthened by the following experiment: Since the  $\alpha$  form transforms to the  $\beta$  form at about 52°C., it should be possible to grow the latter from an alcoholic solution at temperatures greater than 52°C. Accordingly, we tried to precipitate crystals of the "stable"  $\beta$  form from boiling alcohol at 78°C. X-ray photographs proved the precipitate to be the "metastable"  $\alpha$  form, which according to the transformation experiments of Thiessen and Stauff ought not to have been preserved at temperatures above 52°C. The conclusion is inescapable that something in the alcohol is vital to the formation of the  $\alpha$  form, since, if it is exposed to this temperature out of the presence of alcohol, it immediately transforms to the  $\beta$  form. We strongly suspected on several grounds that the two forms differ in water content and that the water is supplied by the alcohol solvent (it is almost impossible to keep even absolute alcohol water-free during the course of an experiment).

3. *Transformation from the  $\alpha$  to the  $\beta$  Form and Its Reversal.*—If the two forms of sodium stearate differ by water content, then it is evident that there is a possibility of passing from either form to the other by the ap-

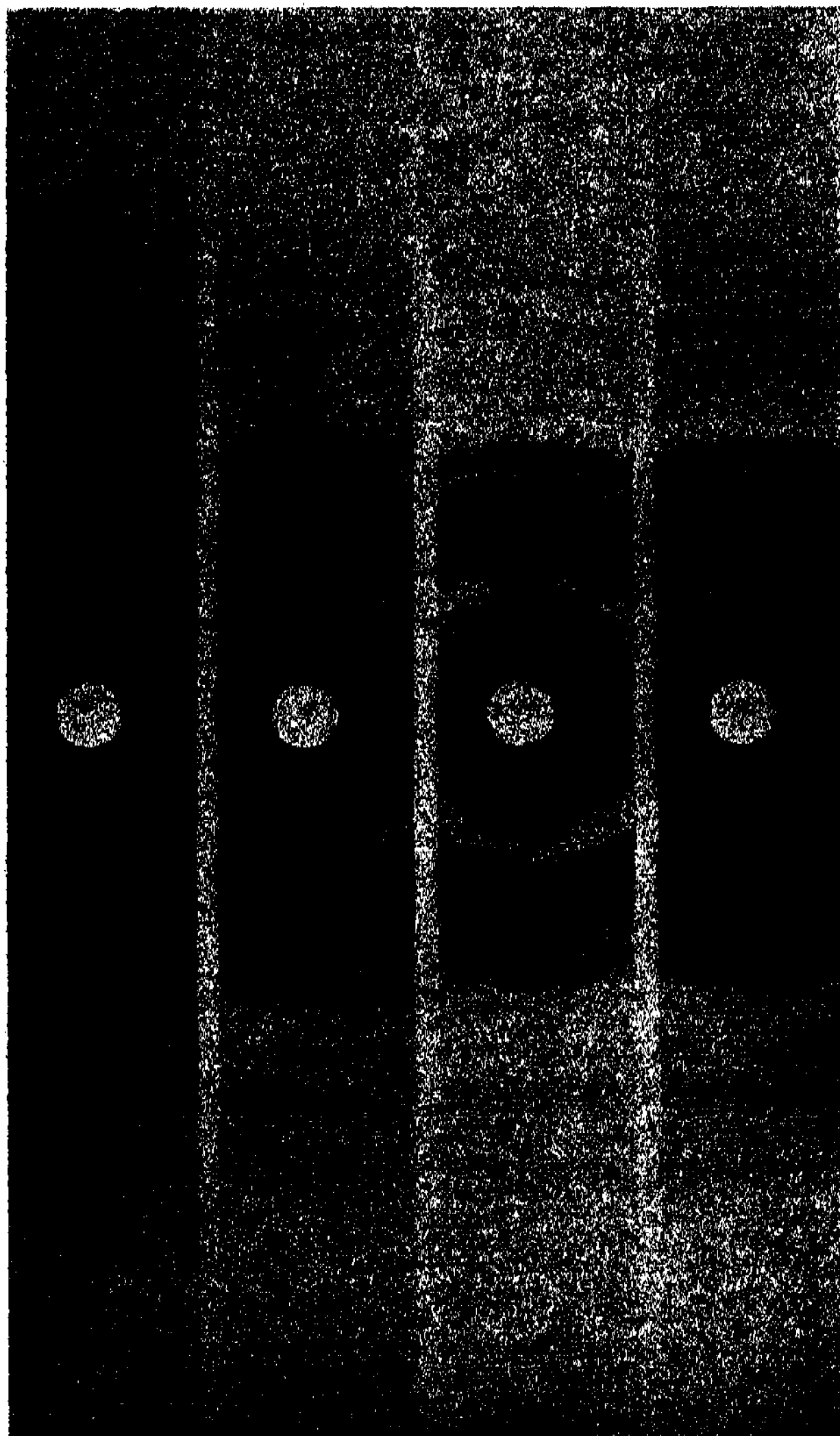


FIGURE 1

X-ray powder photographs ( $\theta$  range, 0 to  $43^\circ$ ) of the following forms of hydrous and anhydrous sodium stearate, left to right:  $\text{NaSt} \cdot \frac{1}{2}\text{H}_2\text{O}$  (" $\alpha$  NaSt");  $\text{NaSt} \cdot \frac{1}{2}\text{H}_2\text{O}$  (" $\beta$  NaSt");  $\text{NaSt}$  (" $\gamma$  NaSt"); form assumed by NaSt after having been heated to any temperature above about  $117^\circ\text{C}$ .

appropriate addition or removal of water. According to our experiments, the removal of water by any of the following means is indeed accompanied

by a transformation from the  $\alpha$  form to the  $\beta$  form, as proved by x-ray powder photographs of the original material and of the product:

- (a) Heating to about  $52^{\circ}\text{C}$ .; a transformation of the  $\alpha$  form to the  $\beta$  form is accompanied by an evolution of water which can be detected on the walls of a tube. This can be done in either an open or closed tube.
- (b) Evacuation of the  $\alpha$  form at room temperature (especially easy with sodium palmitate).
- (c) Grinding at room temperature.

The above evidence might possibly be interpreted as various actions triggering off an inversion from a monotropic to stable form. The following reversal of the transformation cannot be so interpreted, however: If the supposedly stable  $\beta$  form is heated in a sealed tube with a drop of water at  $47^{\circ}\text{C}$ ., the supposedly metastable modification forms at the expense of the supposedly stable one. The only interpretation of this experiment is that a compound of greater hydration is called forth by the presence of the added water.

4. *Transformations between the  $\beta$  Form and  $\gamma$  Form.*—When the  $\beta$  form of sodium stearate is heated, it transforms at about  $103^{\circ}\text{C}$ . to a form called the  $\gamma$  form by de Bretteville.<sup>3, 4</sup> It previously had been assumed that this transformation is a polymorphic inversion. This relation came under suspicion when we noticed that it was prevented by carrying out the heating in a sealed-off tube. This lead us to suspect that the  $\beta$  form was hydrous and that dehydration is prevented in a small closed space. To confirm this, we carried out a transformation cycle with sodium palmitate. The  $\beta$  form was first produced. This was heated in an open tube at  $106^{\circ}\text{C}$ . The x-ray pattern then showed it to have changed into a different form (de Bretteville's  $\gamma$  form). This was then heated with water in a sealed-off tube at either  $100^{\circ}\text{C}$ . or  $200^{\circ}\text{C}$ . In both cases the x-ray pattern of the product was that of the original  $\beta$  form. Subsequent heating in the open at about  $106^{\circ}\text{C}$ . transformed this again into the  $\gamma$  form. This cycle was repeated with the same material as many times as desired. It plainly indicates that water can be added to the  $\gamma$  form to produce the  $\beta$  form and that water can be driven off from the  $\beta$  form to reduce it to the  $\gamma$  form. Thus, the  $\beta$  form is a hydrous form.

5. *Composition of the Several Soap Forms.*—The several pieces of evidence just given establish the fact that the several forms of soap are hydrous in the following order:

- $\alpha$  form (most hydrated)
- $\beta$  form
- $\gamma$  form (least hydrated)

The only remaining question is what is the absolute degree of hydration of each form. Xylol tests established the following hydrate compositions:

$\alpha$  form—sodium stearate  $\frac{1}{2}$  H<sub>2</sub>O  
 $\beta$  form—sodium stearate  $\frac{1}{8}$  H<sub>2</sub>O  
 $\gamma$  form—sodium stearate

6. *X-ray Photographs of Soap Forms.*—In figure 1 we reproduce x-ray powder photographs of the less hydrated forms of soap. In an accompanying paper,<sup>5</sup> x-ray crystallographic data for the hemihydrates are given.

7. *Other Soap Hydrates.*—We have also investigated other hydrates and hope to publish data for them at a later time.

<sup>1</sup> Buerger, M. J., "The General Rôle of Composition in Polymorphism," *Proc. Nat. Acad. Sci.*, **22** 685-689 (1936).

<sup>2</sup> Thiessen, Peter A, and Stauff, Joachim, "Feinbau und Umwandlungen kristallisierter Alkalisalze langkettiger Fettsäuren," *Zeit. Physikal. Chem. (A)*, **176** 397-429 (1936).

<sup>3</sup> Bretteville, A. de, Jr., "X-ray Diffraction Study of Oriented Soaps," Thesis, Stanford University (1942).

<sup>4</sup> Bretteville, Alexander de, Jr., and McBain, J. W., "X-ray Evidence for a Third Polymorphic Form of Sodium Stearate," *Science*, **96**, 470-471 (1942).

<sup>5</sup> Buerger, M. J., "The Characteristics of Soap Hemihydrate Crystals," *Proc. Nat. Acad. Sci.*, **28** 529-535 (1942).

## THE CHARACTERISTICS OF SOAP HEMIHYDRATE CRYSTALS

BY M. J. BUEGER

MINERALOGICAL LABORATORY, MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Communicated October 24, 1942

1. *Introduction.*—In an accompanying paper<sup>1</sup> it is shown that Thiessen and Stauff's<sup>2</sup>  $\alpha$  sodium stearate and  $\alpha$  sodium palmitate are actually the hemihydrates of these compounds. Crystallographic observations on these were also published by Thiessen and Stauff, but their conclusions are in error because the rotating-crystal and powder x-ray diffraction methods, which they employed, are impotent when used to study crystals of large cells and low symmetry. In this paper, the results of a restudy of these crystals by more powerful methods are given.

While a more specific transformation from Thiessen and Stauff's axes is given beyond, the following orientation changes will be of value to those familiar with the literature of soap crystals in understanding the notation used in this paper:

BUEGER'S AXES	AXES CORRESPONDING WITH THIESSEN AND STAUFF'S LATTICE CON- STANTS	AXES CORRESPONDING WITH THIESSEN AND STAUFF'S FIGURE 1
<i>a</i>	<i>b</i>	<i>a</i>
<i>b</i>	<i>a</i>	<i>b</i>
<i>c</i>	<i>c</i>	<i>c</i>

2. *General Crystallographic Characteristics.*—The writer is indebted to both A. de Bretteville, Jr., and F. V. Ryer for experimental work resulting in the production of acceptable single crystals of sodium stearate and sodium palmitate.

The habit of the stearate crystals is well pictured and described by Thiessen and Stauff, although it should be mentioned that their labeling of axes in their figure 1 does not correspond with their labeling of measured cell edges. The crystals are extremely thin tabular parallel to {001} and at the same time elongated parallel to the *b* axis. The tablet is outlined by the pinacoid {100} and the prism {110}.

The habit of the palmitate crystals as grown from the alcohol solutions containing grains of sodium chloride is also very thin tabular parallel to {001}, but the outline of the tablet is diamond-shaped due to the exclusive development of {110}; the form {100} present in the stearate is missing in the palmitate. This variation in habit presumably is due to growth of the palmitate in the presence of an impurity, sodium chloride.

The following further observations were made specifically on the stearate, but apply, at least in part, to the palmitate also: The crystals have eminent {001} cleavage. They also have good {010} cleavage, although this is ordinarily overshadowed by the {001} cleavage. The plates are extremely plastic due to easy translation gliding with  $T = \{001\}$ ,  $t =$  any direction in this plane, but [010] preferred; crystals bend easily, with a special tendency to bend with [100] as folding axis.

There is a structural relation between the possibility of bending crystals about a folding axis and the growth of crystals bent about the folding axis. A common kind of stearate crystal is an arborescent growth in which the main trunk crystal splits up, each split branching away from it by bending about the folding axis. These branches commonly bend away from the trunk, then bend back parallel with it, all the bending taking place about the folding axis [100]. In a crop of crystals grown from an alcoholic solution which has been allowed to stand some time, it is not uncommon to find a complexly branched individual occupying several square centimeters in total exposed bent and branched area, yet only as thick as the width of the trunk, a small fraction of a millimeter.

3. *Symmetry.*—In view of the difference between the results reported here and those previously reported by Thiessen and Stauff, they were checked by several methods.



For sodium stearate, equi-inclination Weissenberg photographs were taken of zero, first and second levels for rotation about the axis of elongation (Fig. 4). A set of equal-cone de Jong-Bouman photographs (Figs. 1 and 2) of the zero, first, second, third and fourth levels for the same rotation axis were also made.

All these photographs clearly indicate a level symmetry of  $C_2$  (see Figs. 1, 2 and 4). This constitutes sufficient proof that the centrosymmetrical symmetry of the crystal is  $C_{2h}$ ,  $2/m$ ; consequently the crystal is monoclinic. This symmetry was established with sufficient data for the palmitate also. The axis of elongation of the stearate is consequently the  $b$  axis.

4. *Space Lattice Type*.—The photographs just discussed show that the  $(010)^*$  planes are based upon a parallelogram net, but the reflections recorded by both Weissenberg and de Jong-Bouman photographs are so close together in the direction of the vector  $c^*$  that it is extremely difficult to be sure exactly how adjacent levels of the reciprocal lattice superpose. In order to resolve this difficulty and determine the correct stacking sequence, recourse was had to the moving film precession method<sup>8</sup> with  $[100]$  as the generator of the precession cone. This showed that the  $(100)^*$  planes are based upon a diamond net and that these nets in adjacent levels are exactly superposed by the  $a^*$  translation. This information, taken with that given above, uniquely determines the lattice as  $A$ -centered monoclinic.

5. *Space Group*.—The only doubled translations in the reciprocal lattice are all  $a^*$  in central  $(010)^*$ . This indicates a glide  $a$ . The diffraction symbol of the crystal is therefore  $2/mA - /a$ . This is consistent with either of the two space groups  $Aa$  or  $A2/a$ .

6. *Cell Dimensions*.—The following cell dimensions were determined from the  $b$ -axis rotation photograph and various de Jong-Bouman and precession photographs (accuracy estimated at about  $1/2\%$ ):

	NA STEARATE $\cdot 1/2 H_2O$	NA PALMITATE $\cdot 1/2 H_2O$
$a$	9.16 A	9.13 A
$b$	8.00	8.01
$c$	103.96	91.85
$\beta$	$93^\circ 43'$	...

The matrix of the transformation from the axes given by Thiessen and Stauff to those of Buerger is

$$\begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{vmatrix}$$



These crystals contain 16 hemihydrated molecules per unit cell. It is interesting to note that, if the crystals were anhydrous, as formerly sup-

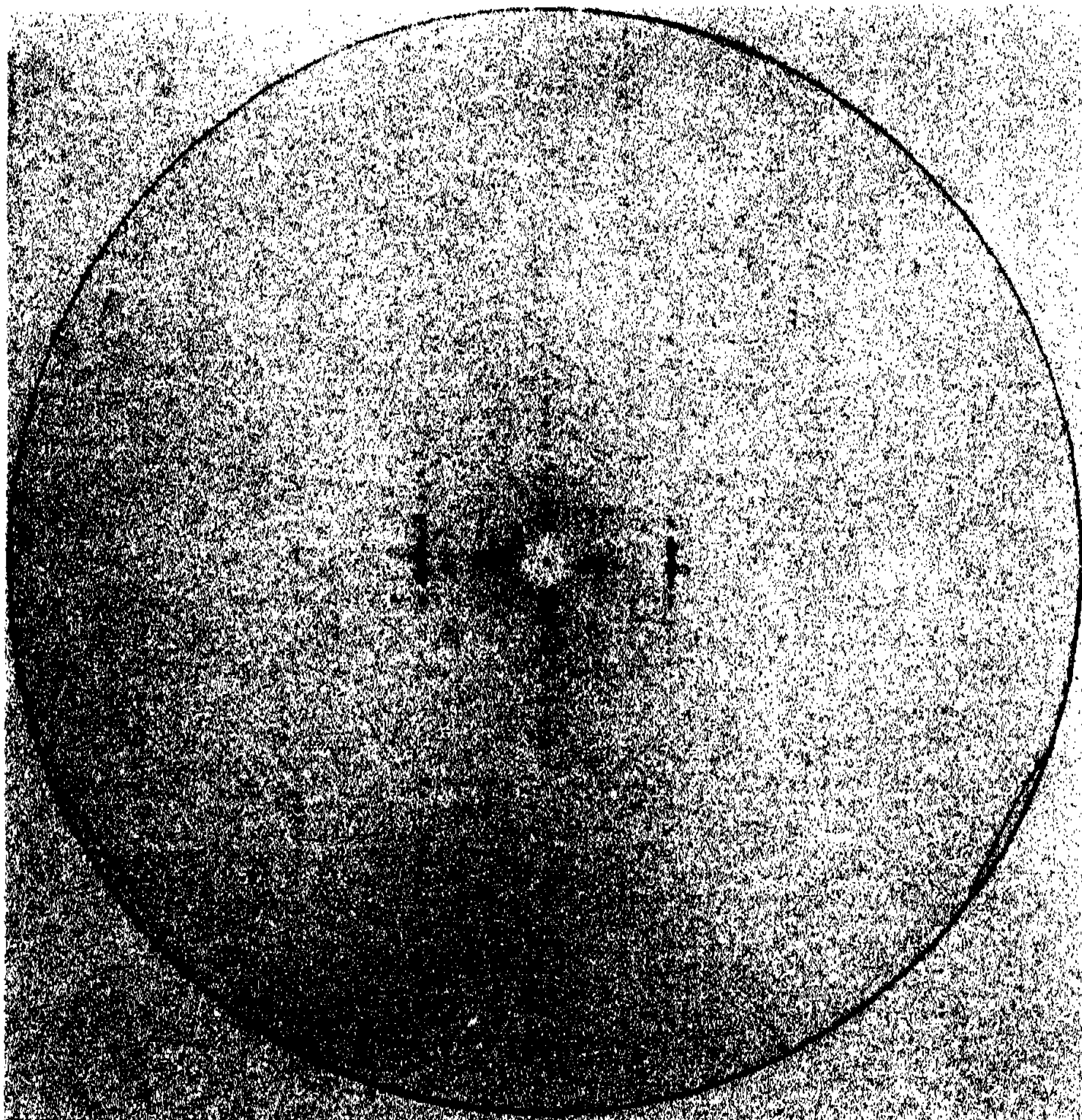


FIGURE 1

de Jong-Bouman photograph,  $\text{NaSt} \cdot \frac{1}{2}\text{H}_2\text{O}$ ; zero level (010)\* net. ( $\text{CuK}\alpha$  radiation.)

posed, this would require 16 chemically identical molecules in a unit cell whose space group provides, at most, an 8-fold position, a somewhat difficult situation to understand. The hemihydrate character, however, is equivalent to having to accommodate only eight groups of the type  $2\text{NaSt} \cdot \text{H}_2\text{O}$  in the cell, and the eight water molecules exactly fill the eight

general equipoints of the one of the two space group possibilities,  $A2/a$ . The alternative space group,  $Aa$ , is unlikely because it can provide only four equipoints for the eight waters.

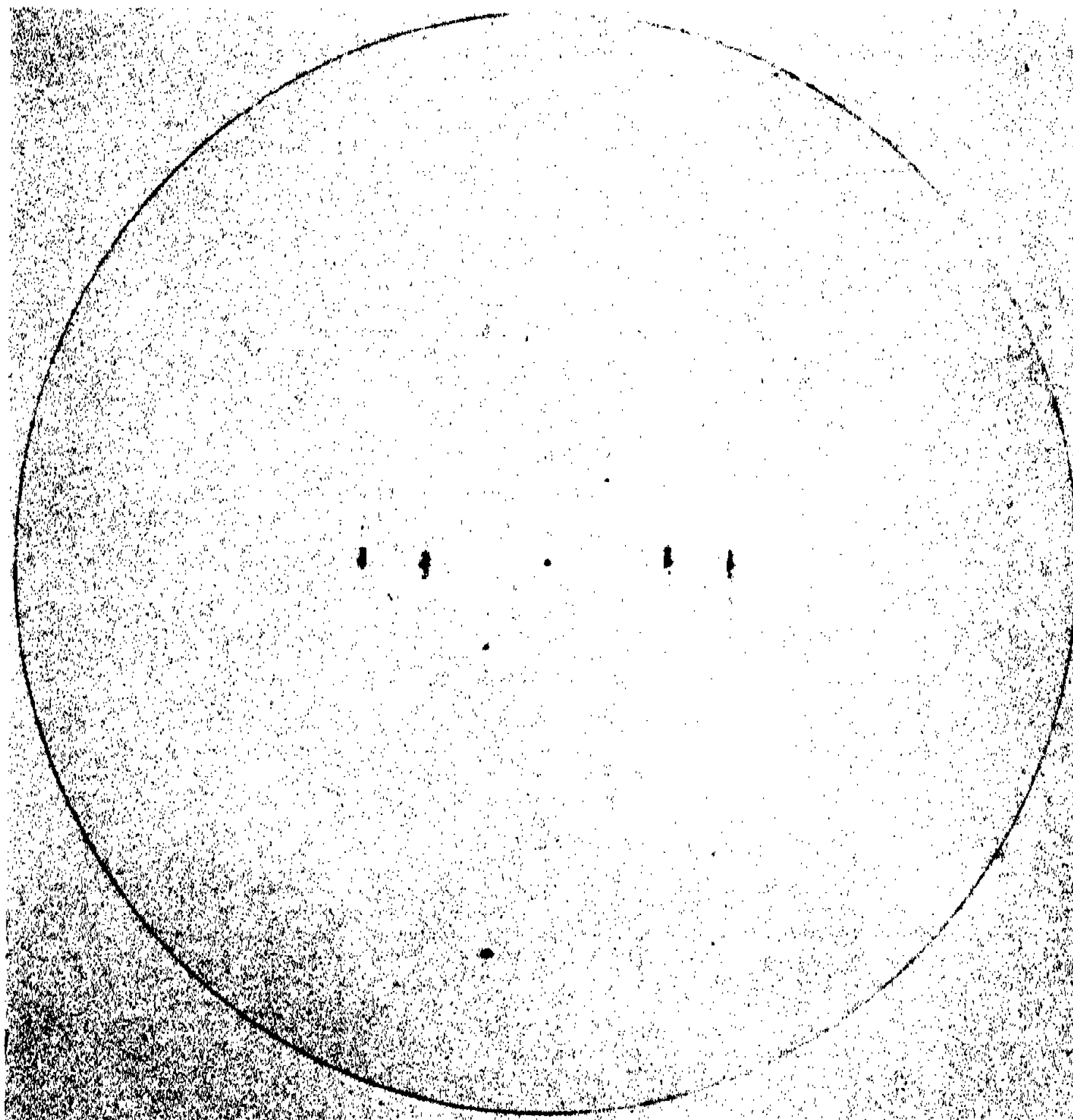


FIGURE 2

de Jong-Bouman photograph,  $\text{NaSt} \cdot \frac{1}{2}\text{H}_2\text{O}$ ; second level (010)\* net. ( $\text{CuK}\alpha$  radiation.)

7. *Indexing Powder Photographs of Soap Crystals.*—There are those who would attempt to assign indices to the lines of powder photographs of soap crystals. Even a casual comparison of Weissenberg, rotating crystal and powder photographs of any soap crystal makes it most apparent that each line on the powder photograph is actually a band composed of many re-



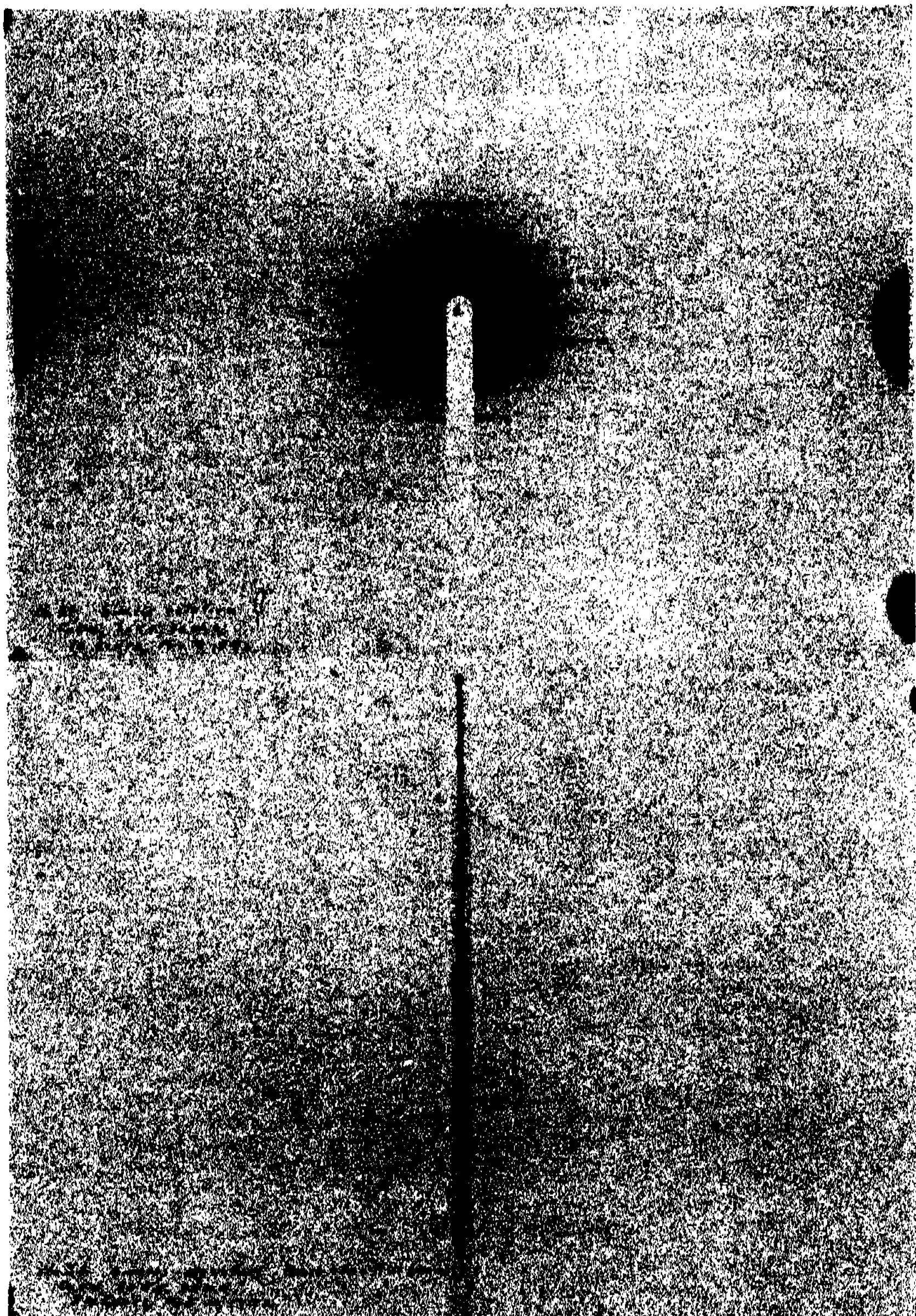


Figure 3 (above). Rotating-crystal photograph,  $\text{NaSt} \cdot \frac{1}{2}\text{H}_2\text{O}$ ;  $b$  axis rotation.

Figure 4 (below). Weissenberg photograph,  $\text{NaSt} \cdot \frac{1}{2}\text{H}_2\text{O}$ ;  $b$  axis rotation, zero level. Note the sets of spots nearly equidistant from the center line, especially near the bottoms of festoons. In the rotating-crystal photograph above, the spots of a set are unresolved and consequently the rotating-crystal photograph cannot be indexed.

flections of nearly identical  $\sin \theta$  value. Any indexing of powder photographs of soap crystals is consequently very questionable. It hardly need be said that no notion of symmetry can be obtained from powder photographs of crystals with such large cells as soaps.

There is also a prevailing opinion in some quarters that if the computed length of a soap molecule fits the spacing  $d_{(001)}$ , the crystal is orthorhombic, while if it must be tipped to fit the spacing, the crystal is monoclinic. Neither of these propositions is necessarily true.

<sup>1</sup> Buerger, M. J., Smith, L. B., Bretteville, A. de, Jr., and Ryer, F. V., "The Lower Hydrates of Soap," *Proc. Nat. Acad. Sci.*, 28, 526-529 (1942).

<sup>2</sup> Thiessen, Peter A., and Stauff, Joachim, "Feinbau und Umwandlungen kristallisierter Alkalisalze langkettiger Fettsäuren," *Zeit. Physikal. Chem. (A)*, 176, 397-429 (1936).

<sup>3</sup> Buerger, M. J., "X-ray Crystallography," 210-211, John Wiley & Sons, Inc., New York, 1942.

## STATISTICAL METRICS

BY KARL MENGER

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF NOTRE DAME

Communicated October 27, 1942

We shall call *statistical metric* a set  $S$  such that with each two elements ("points")  $p$  and  $q$  of  $S$  a probability function  $\Pi(x; p, q)$  is associated satisfying the following conditions:

1.  $\Pi(0; p, p) = 1$ .
2. If  $p \neq q$ , then  $\Pi(0; p, q) < 1$ .
3.  $\Pi(x; p, q) = \Pi(x; q, p)$ .
4.  $T[\Pi(x; p, q), \Pi(y; q, r)] \leq \Pi(x + y; p, r)$ .

where  $T(\alpha, \beta)$  is a function defined for  $0 \leq \alpha \leq 1$  and  $0 \leq \beta \leq 1$  such that

- (a)  $0 \leq T(\alpha, \beta) \leq 1$ .
- (b)  $T$  is non-decreasing in either variable.
- (c)  $T(\alpha, \beta) = T(\beta, \alpha)$ .
- (d)  $T(1, 1) = 1$ .
- (e) If  $\alpha > 0$ , then  $T(\alpha, 1) > 0$ .

By a probability function we mean a non-decreasing function defined for all non-negative values of  $x$ , continuous to the right, with values between 0 and 1, and converging toward 1 as  $x$  increases beyond all bounds.

We call  $\Pi(x; p, q)$  the *distance function* of  $p$  and  $q$  and interpret it as the probability that the points  $p$  and  $q$  have a distance  $\leq x$ . Condition 4,

our "triangular inequality," implies that  $\Pi(z; p, r) \geq \text{Max. } T[\Pi(x; p, q), \Pi(z - x; q, r)]$  for all points  $q$  and all numbers  $x$  between 0 and  $z$ . We shall call the function  $T$  the *triangular norm* of the statistical metric, and more specifically refer to the metric defined above as a  $T$ -metric. A triangular norm  $T$  will be called *simple* if

$$(f) \quad 0 < T(\alpha, \beta) < 1 \text{ for } 0 < \alpha \cdot \beta < 1.$$

An ordinary metric space is a statistical metric such that for each pair of points  $p, q$  there exists a number  $d(p, q) \geq 0$  with the property that  $\Pi(x; p, q) = 0$  if  $x < d(p, q)$ , and  $= 1$  if  $x \geq d(p, q)$ .

On the basis of our postulates large parts of metric geometry can be developed, in particular, a theory of betweenness. We shall say that  $q$  lies *between*  $p$  and  $r$  (and we write  $pqr$ ) if

$$T[1 - \Pi(x; p, q), 1 - \Pi(y; q, r)] \leq 1 - \Pi(x + y; p, r).$$

Equivalent is the assumption

$$\Pi(z; p, r) \leq 1 - \text{Max. } T[1 - \Pi(x; p, q), 1 - \Pi(z - x; q, r)] \text{ for } 0 \leq x \leq z.$$

Obviously, if  $pqr$  then  $rqp$ . In metric spaces if  $q$  and  $r$  are distinct, then  $pqr$  and  $prq$  are incompatible. In a statistical metric we can only prove: If  $q$  and  $r$  are apart, then  $pqr$  and  $prq$  are incompatible, where  $q$  and  $r$  are said to be apart if there exists a number  $y > 0$  such that  $\Pi(y; q, r) = 0$ .

If for each two points of a statistical metric  $S$  the distance function  $\Pi(x; p, q)$  belongs to a family  $\mathfrak{P}$  of probability functions, we call  $S$  metrized by means of  $\mathfrak{P}$ . Let  $\mathfrak{P}$  be a 2 parameter family of probability functions  $\Pi(x; a', a'')$  defined for all real numbers  $a'$  and  $a''$  such that  $0 \leq a' \leq a''$  and satisfying the conditions

$$\begin{aligned} \Pi(x; a', a'') &= 0 \text{ if } x \leq a', \\ \Pi(x; a', a'') &= 1 \text{ if } x \geq a'', \\ 0 < \Pi(x; a', a'') < 1 &\text{ for } a' < x < a''. \end{aligned}$$

If  $S$  is metrized by means of  $\mathfrak{P}$  and  $T$  is simple, then for each three points  $p, q, r$  one of which lies between the two other ones, two of the distance functions  $\Pi(x; p, q), \Pi(y; q, r), \Pi(z; p, r)$  determine the third. In particular, if  $pqr$  and  $\Pi(x; p, q) = \Pi(x; a', a'')$  and  $\Pi(y; q, r) = \Pi(y; b', b'')$ , then  $\Pi(z; p, r) = \Pi(z; a' + b', a' + b'')$ . From this theorem one readily derives the classical law: If  $pqr$  and  $prs$ , then  $pqs$  and  $qrs$ .

If

$$\begin{aligned} \Pi(x; p, q) &= \Pi(x; r, s) = \Pi(x; a', a'') \\ \Pi(x; q, r) &= \Pi(x; p, s) = \Pi(x; b', b'') \\ \Pi(x; p, r) &= \Pi(x; q, s) = \Pi(x; a' + b', a' + b''), \end{aligned}$$

then  $p, q, r, s$  form what may be called a *pseudo-linear statistical quadruple*, i.e., a quadruple which cannot be ordered by means of the between-relation though for each three of the four points one lies between the other two.

If a statistical T-metric  $S$  metrized by means of  $\mathbb{B}$  contains more than four points, then by virtue of the properties of betweenness this relation can be used to order  $S$ . Moreover, the other ideas of metric geometry (convexity, geodesics, etc.) can be applied.

The three principal applications of statistical metrics are to macroscopic, microscopic and physiological spatial measurements. Statistical metrics are designed to provide us (1) with a method removing conceptual difficulties from microscopic physics and transferring them into the underlying geometry, (2) with a treatment of thresholds of spatial sensation eliminating the intrinsic paradoxes of the classical theory. For a given point  $p_0$  the number  $\Pi(0; p_0, q)$  considered as a function of the point  $q$  indicates the probability that  $q$  cannot be distinguished from  $p_0$ . The study of this function should replace the attempt to determine a definite set of points  $q$  which cannot be distinguished from  $p_0$ . This function could also be used advantageously instead of a relation of physical identity for which, as Poincaré emphasized on several occasions, we always have triples  $p, q, r$  for which

$$p = q, q = r, \text{ and } p \neq r.$$

Experiments indicate that  $q$  sometimes can and sometimes cannot be distinguished from  $p_0$ . Hence, the adequate description of the situation seems to arise from counting the relative frequency of these occurrences.

## *NATURAL ISOMORPHISMS IN GROUP THEORY*

BY SAMUEL EILENBERG AND SAUNDERS MACLANE

DEPARTMENTS OF MATHEMATICS, UNIVERSITY OF MICHIGAN AND HARVARD UNIVERSITY

Communicated October 26, 1942

*1. Introduction.*—Frequently in modern mathematics there occur phenomena of “naturalness”: a “natural” isomorphism between two groups or between two complexes, a “natural” homeomorphism of two spaces and the like. We here propose a precise definition of the “naturalness” of such correspondences, as a basis for an appropriate general theory. In this preliminary report we restrict ourselves to the natural isomorphisms of group theory; with this limitation we can present the basic concepts of our theory without developing the axiomatic approach necessary for a general treatment applicable to various branches of mathematics.

Properties of character groups (see the definitions in § 5 below) may serve to illustrate the ideas involved. Thus, it is often asserted that the character group of a finite group  $G$  is isomorphic to the group itself, but not in a "natural" way. Specifically, if  $G$  is cyclic of prime order  $p$ , there is for each generator of  $G$  an isomorphism of  $G$  to its character group, so that the proof furnishes  $p - 1$  such isomorphisms, no one of which is in any way distinguished from its fellows. However, the proof that the character group of the character group of  $G$  is isomorphic to  $G$  itself is considered "natural," because it furnishes for each  $G$  a unique isomorphism, not dependent on any choice of generators.

To give these statements a clear mathematical meaning, we shall regard the character group  $Ch(G)$  of  $G$  as a function of a variable group  $G$ , together with a prescription which assigns to any homomorphism  $\gamma$  of  $G$  into a second group  $G'$ ,

$$\gamma: G \rightarrow G',$$

the induced homomorphism (see (5) below)

$$Ch(\gamma): Ch(G') \rightarrow Ch(G).$$

The functions  $Ch(G)$  and  $Ch(\gamma)$  jointly form what we shall call a "functor"; in this case, a "contravariant" one, because the mapping  $Ch(\gamma)$  works in a direction opposite to that of  $\gamma$ . A natural isomorphism between two functions of groups will be an isomorphism which commutes properly with the induced mappings of the functors.

With our description of a natural isomorphism, practically all the general isomorphisms obtained in group theory and its applications (homology theory, Galois theory, etc.) can be shown to be "natural." This results in added clarity in such situations. Furthermore, there are definite proofs where the naturality of an isomorphism is needed, especially when a passage to the limit is involved. In fact, our condition (E2) below appears in the definition of the isomorphism of two direct or two inverse systems of groups.<sup>1</sup>

2. *Functors*.—The definition of a functor will be given for the typical case of a functor  $T$  which depends on two groups as arguments, and is *covariant* in the first argument and *contravariant* in the second. Such a functor is determined by two functions. The *group function* determines for each pair of topological groups  $G$  and  $H$  (contained in a given legitimate set of groups) another group  $T(G, H)$ . The *mapping function* determines for each pair of homomorphisms<sup>2</sup>  $\gamma: G_1 \rightarrow G_2$  and  $\eta: H_1 \rightarrow H_2$  a homomorphism  $T(\gamma, \eta)$ , such that

$$T(\gamma, \eta): T(G_1, H_2) \rightarrow T(G_2, H_1). \quad (1)$$



We require that  $T(\gamma, \eta)$  be the identity isomorphism whenever  $\gamma$  and  $\eta$  are identities, and that, whenever the products  $\gamma_2\gamma_1$  and  $\eta_2\eta_1$  are defined,

$$T(\gamma_2\gamma_1, \eta_2\eta_1) = T(\gamma_2, \eta_1)T(\gamma_1, \eta_2). \quad (2)$$

Some functors will be defined only for special types of groups (e.g., for abelian groups) or for special types of homomorphisms (e.g., for homomorphisms "onto").

If  $\gamma$  and  $\eta$  are both isomorphisms,<sup>3</sup> it follows from these conditions that  $T(\gamma, \eta)$  is also an isomorphism. Consequently, if the groups  $G_1$  and  $G_2$  and the groups  $H_1$  and  $H_2$  are isomorphic, the functor  $T$  gives rise to isomorphic groups  $T(G_1, H_1)$  and  $T(G_2, H_2)$ .

3. *Examples.*—The direct product  $G \times H$  of two groups may be regarded as the group function of a functor. The corresponding mapping function specifies, for each pair of homomorphisms  $\gamma: G_1 \rightarrow G_2$  and  $\eta: H_1 \rightarrow H_2$ , an induced homomorphism  $\gamma \times \eta$ , defined for every element  $(g_1, h_1)$  in  $G_1 \times H_1$  as

$$[\gamma \times \eta](g_1, h_1) = (\gamma g_1, \eta h_1).$$

Then

$$\gamma \times \eta: G_1 \times H_1 \rightarrow G_2 \times H_2, \quad (3)$$

and, whenever  $\gamma_2\gamma_1$  and  $\eta_2\eta_1$  are defined, one has

$$(\gamma_2\gamma_1) \times (\eta_2\eta_1) = (\gamma_2 \times \eta_2)(\gamma_1 \times \eta_1). \quad (4)$$

Except for the absence of contravariance, these conditions are parallel to (1) and (2), hence  $G \times H$ ,  $\gamma \times \eta$  define a functor, covariant in both  $G$  and  $H$ .

Whitney's tensor product<sup>4</sup>  $G \circ H$  of two discrete groups<sup>5</sup>  $G$  and  $H$  is the group function of a functor. The elements of this group are all finite sums  $\sum g_i \circ h_i$  of formal products  $g_i \circ h_i$ ; the group operation is the obvious addition, and the relations are  $g \circ (h + h') = g \circ h + g \circ h'$  ( $g + g' \circ h = g \circ h + g' \circ h$ ). Given two homomorphisms  $\gamma: G_1 \rightarrow G_2$  and  $\eta: H_1 \rightarrow H_2$ , there is an induced homomorphism  $\gamma \circ \eta$  of  $G_1 \circ H_1$  into  $G_2 \circ H_2$ , defined for any generator  $g_1 \circ h_1$  of  $G_1 \circ H_1$  as

$$[\gamma \circ \eta](g_1 \circ h_1) = (\gamma g_1) \circ (\eta h_1) \in G_2 \circ H_2.$$

Formulae (3) and (4), with the cross replaced by the circle, again hold, so that  $G \circ H$ ,  $\gamma \circ \eta$  determine a functor of discrete groups, covariant in both arguments.

In a similar fashion, the free product of two groups leads to a functor.

An important functor is given by the group of all homomorphisms  $\phi$  of a fixed locally compact topological abelian group  $G$  into another topological abelian group  $H$ . The sum of two such homomorphisms  $\phi_1$  and  $\phi_2$  is de-



defined for each  $g \in G$  by setting  $(\phi_1 + \phi_2)(g) = \phi_1(g) + \phi_2(g)$ . Under this operation, all  $\phi: G \rightarrow H$  constitute a group  $\text{Hom}(G, H)$ ; it carries an appropriate topology, the description of which we omit. For given  $\gamma: G_1 \rightarrow G_2$  and  $\eta: H_1 \rightarrow H_2$  and for each  $\phi \in \text{Hom}(G_2, H_1)$  we have

$$\begin{array}{ccccc} & \gamma & \phi & \eta & \\ & \downarrow & \downarrow & \downarrow & \\ G_1 & \rightarrow & G_2 & \rightarrow & H_1 \rightarrow H_2. \end{array}$$

Consequently we define  $\text{Hom}(\gamma, \eta)(\phi) = \eta\phi\gamma$ , and verify that

$$\text{Hom}(\gamma, \eta): \text{Hom}(G_2, H_1) \rightarrow \text{Hom}(G_1, H_2),$$

$$\text{Hom}(\gamma_2\gamma_1, \eta_2\eta_1) = \text{Hom}(\gamma_1, \eta_2) \text{Hom}(\gamma_2, \eta_1).$$

Clearly when  $\gamma$  and  $\eta$  are identity mappings of  $G$  and  $H$  the induced mapping  $\text{Hom}(\gamma, \eta)$  is the identity mapping of  $\text{Hom}(G, H)$  on itself. Hence the functions  $\text{Hom}(G, H)$  and  $\text{Hom}(\gamma, \eta)$  determine for abelian groups a functor  $\text{Hom}$ , covariant in  $H$  and contravariant in  $G$ .

The special case when  $H$  is the group  $P$  of reals modulo 1 furnishes the character group,

$$\text{Ch}(G) = \text{Hom}(G, P), \quad \text{Ch}(\gamma) = \text{Hom}(\gamma, e)$$

where  $e$  is the identity mapping of  $P$  on itself. Therefore the character group is a contravariant functor, defined for abelian groups. Explicitly, if we express the result  $\chi(g)$  of applying the character  $\chi$  to the element  $g \in G$  as the value (a real number modulo 1) of the bilinear form  $(g, \chi)$ , the definition of  $\text{Ch}(\gamma)$  can be written as

$$(g, \text{Ch}(\gamma)\chi') = (\gamma g, \chi'), \quad g \in G, \quad \chi' \in \text{Ch}(G'). \quad (5)$$

4. *Equivalence of Functors.*—Let  $T$  and  $S$  be two functors which are, say, both covariant in the variable  $G$  and contravariant in  $H$ . Suppose that for each pair of groups  $G$  and  $H$  we are given a homomorphism

$$\tau(G, H): T(G, H) \rightarrow S(G, H).$$

We say that  $\tau$  establishes a natural *equivalence* of the functor  $T$  to the functor  $S$  and that  $T$  is naturally *equivalent* to  $S$  (in symbols,  $\tau: T \longleftrightarrow S$ ) whenever

(E1) Each  $\tau(G, H)$  is a bicontinuous isomorphism of  $T(G, H)$  onto  $S(G, H)$ ;

(E2) For each  $\gamma: G_1 \rightarrow G_2$  and  $\eta: H_1 \rightarrow H_2$ ,  
 $\tau(G_2, H_1)T(\gamma, \eta) = S(\gamma, \eta)\tau(G_1, H_2).$

The first requirement insures the term-by-term isomorphism of the two group functions  $T(G, H)$  and  $S(G, H)$ , while the second requirement is

precisely the "naturality" condition. It can be shown that the condition (E2) is implied by two special cases; the case when  $\eta$  is an identity, and the case when  $\gamma$  is an identity.

This relation of natural equivalence between functors is reflexive, symmetric and transitive. In many cases we dispense with condition (E1), and obtain a more general concept of a "transformation" of a functor  $T$  into a functor  $S$ .

5. *Examples of Natural Equivalence.*—The well-known isomorphism

$$G \cong Ch(Ch(G)) \quad (6)$$

for locally compact abelian groups, can be regarded as an equivalence of functors, and is in this sense *natural*. The right-hand side of (6) suggests the covariant functor,  $Ch^2$ , defined by iteration of the functor  $Ch$ , as

$$Ch^2(G) = Ch(Ch(G)), \quad Ch^2(\gamma) = Ch(Ch(\gamma)).$$

The left-hand side of (6) suggests the identity functor,  $I$ ,

$$I(G) = G, \quad I(\gamma) = \gamma.$$

The bilinear form  $(g, \chi) = \chi(g)$  determines to each character  $\chi \in Ch(G)$  and each  $g \in G$  a real number modulo 1; similarly the form  $(\chi, h) = h(\chi)$  is defined for each  $h \in Ch^2(G)$ . The form  $(g, \chi)$ , regarded as a function of  $\chi$  for fixed  $g$ , is a character  $h$  in  $Ch^2(G)$  which we call  $[\tau(G)]g$ . Explicitly, this definition of  $\tau$  reads

$$(\chi, \tau(G)g) = (g, \chi), \quad g \in G, \quad \chi \in Ch(G).$$

The validity of condition (E1) for  $\tau(G)$  is the basic theorem of character theory. The condition (E2) asserts that in the diagram

$$\begin{array}{ccc} & \tau(G) & \\ G & \xrightarrow{\quad} & Ch^2(G) \\ \downarrow \gamma & & \downarrow Ch^2(\gamma) \\ & \tau(G') & \\ G' & \xrightarrow{\quad} & Ch^2(G') \end{array}$$

the two paths leading from  $G$  to  $Ch^2(G')$  have the same effect, or that, for each  $g \in G$ , both elements  $\tau(G')\gamma g$  and  $Ch^2(\gamma)\tau(G)g$  are identical as elements of  $Ch^2(G')$ . This means that, for each  $\chi \in Ch(G')$ , one should have

$$(\chi', \tau(G')\gamma g) = (\chi', Ch^2(\gamma)\tau(G)g).$$

By the definition of  $\tau$ , the expression on the left is simply  $(\gamma g, \chi')$ . By successive application to the expression on the right of the definitions of  $Ch$ ,  $\tau$  and  $Ch$ , we obtain

$$(\chi', Ch^2(\gamma)\tau(G)g) = (Ch(\gamma)\chi', \tau(G)g) = (g, Ch(\gamma)\chi') = (\gamma g, \chi').$$

The identity of these results shows that we do have a natural equivalence  $\tau(G): G \longleftrightarrow Ch^2(G)$ .

When  $G$  is finite, the isomorphism  $G \rightarrow Ch G$  cannot be "natural" according to our definitions, for the simple reason that the functor  $I$  on the left is covariant, while the functor  $Ch$  on the right is contravariant.

As other examples of equivalences between functors, we may cite the usual isomorphisms which give the associative and commutative laws for the direct product, the tensor product and the free product. Various distributive laws, such as

$$(G_1 \times G_2) \circ H \cong (G_1 \circ H) \times (G_2 \circ H),$$

$$Hom(G_1 \times G_2, H) \cong Hom(G_1, H) \times Hom(G_2, H),$$

when established with the obvious isomorphisms, are in fact equivalences between functors.

A less obvious relation between the tensor product and the functor " $Hom$ " is<sup>6</sup>

$$Hom(G, Hom(H, K)) \cong Hom(G \circ H, K), \quad (7)$$

where  $G$  and  $H$  are discrete abelian groups,  $K$  a topological abelian group. This isomorphism is obtained by a correspondence  $\tau(G, H, K)$  which specifies for each element  $\phi \in Hom(G, Hom(H, K))$  a corresponding homomorphism in  $Hom(G \circ H, K)$ , defined for any generator  $g \circ h$  of  $G \circ H$  as

$$[\tau(G, H, K)](\phi)(g \circ h) = [\phi(g)](h) \text{ in } K.$$

One may show that  $\tau$  does give an isomorphism, bicontinuous in the appropriate topologies. Both sides of (7) may be treated as the group functions of functors which are obtained by composition from " $Hom$ " and " $\circ$ ." The corresponding mapping functions, for given homomorphisms

$$\gamma: G_1 \rightarrow G_2, \quad \eta: H_1 \rightarrow H_2, \quad \kappa: K_1 \rightarrow K_2,$$

are defined by a parallel composition as

$$Hom(\gamma, Hom(\eta, \kappa)), \quad Hom(\gamma \circ \eta, \kappa).$$

Both functors are contravariant in  $G$  and  $H$ , covariant in  $K$ .

The naturality condition for the isomorphism  $\tau$  reads

$$\tau(G_1, H_1, K_2) Hom(\gamma, Hom(\eta, \kappa)) = Hom(\gamma \circ \eta, \kappa) \tau(G_2, H_2, K_1).$$

Both sides, when applied to an element  $\phi \in Hom(G_2, Hom(H_2, K_1))$  yield a homomorphism in  $Hom(G_1 \circ H_1, K_2)$ . If each of these homomorphisms is applied to a typical generator  $g_1 \circ h_1$  of the tensor product  $G_1 \circ H_1$ , straightforward application of the relevant definitions shows that the same element of  $K_2$  is obtained in both cases; namely,  $\kappa\{[\phi(\gamma(g_1))](\eta(h_1))\}$ .

One may also see directly that this expression represents the only way of constructing an element of  $K_2$  from the elements  $g_1$  and  $h_1$  and the mappings  $\kappa$ ,  $\phi$ ,  $\gamma$  and  $\eta$ .

The natural isomorphism (7) has some interesting consequences. If  $K$  is taken to be the group  $P$  of real numbers modulo 1,  $\text{Hom}(H, K)$  becomes the character group  $\text{Ch}(H)$ , and the formula may be written as

$$\text{Hom}(G, \text{Ch } H) \cong \text{Ch}(G \circ H).$$

Applying the functor  $\text{Ch}$  to both sides and using the natural equivalence of  $\text{Ch}^2$  and  $I$ , we obtain the equivalence

$$G \circ H \cong \text{Ch } \text{Hom}(G, \text{Ch } H).$$

Since this is "natural," this could be used as a definition of the tensor product  $G \circ H$ .

6. *Generalizations.*—With the appropriate definition of a normal subfunctor  $S$  of a functor  $T$  one can construct a quotient functor  $T/S$ , whose group function has as its values quotient groups (i.e., factor groups). With this operation, all the standard constructions on groups may be represented as group functions of suitable functors.

An inspection of the concept of a functor and of a natural equivalence shows that they may be applied not only to groups with their homomorphisms, but also to topological spaces with their continuous mappings, to simplicial complexes with their simplicial transformations, and to Banach spaces with their linear transformations. These and similar applications can all be embodied in a suitable axiomatic theory. The resulting much wider concept of naturality, as an equivalence between functors, will be studied in a subsequent paper.

<sup>1</sup> Pontrjagin, L., "Ueber den algebraischen Inhalt der topologische Dualitätssätze," *Mathematische Ann.*, 105, 165–205 (1931). Lefschetz, S., "Algebraic Topology," *Am. Math. Soc. Colloquium Pub.*, 27, 55 (1942).

<sup>2</sup> By a homomorphism we mean a *definite* pair of groups  $G_1$  and  $G_2$  and a (continuous) homomorphic mapping  $\gamma_1$  of the first onto a subgroup of the second. The product  $\gamma_2\gamma_1$  is defined for those pairs  $\gamma_1: G_1 \rightarrow G_2$ ,  $\gamma_2: G_1' \rightarrow G_2$  with  $G_2 = G_1'$ .

<sup>3</sup> By an isomorphism we mean a homomorphism of  $G_1$  onto  $G_2$  which is one-one and bicontinuous.

<sup>4</sup> Whitney, H., "Tensor Products of Abelian Groups," *Duke Math. Jour.*, 4, 495–528 (1938).

<sup>5</sup> Here and subsequently the group operation in  $G$  and in  $H$  is written as addition, whether or not the groups are abelian.

<sup>6</sup> This isomorphism was established by the authors; cf. *Ann. Math.*, 44 (1943).

## CONCERNING INTERSECTING CONTINUA

BY R. L. MOORE

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF TEXAS

Communicated November 2, 1942

In this paper a study is made of certain relationships involving (1) a continuum  $M$ , (2) its boundary  $\beta$ , (3) the components of  $M - \beta$ , (4) the components of the complements of the closures of the components of  $M - \beta$  and (5) the common part of  $M$  and another continuum  $K$ . Numbered axioms and chapters referred to are axioms and chapters of the author's book, "Foundations of Point Set Theory."<sup>1</sup>

**THEOREM 1.** *If, in a space satisfying Axioms 0 and 1,  $M$  is a compact continuum, with a boundary  $\beta$  such that, if  $D$  is a component of  $M - \beta$ ,  $\beta \cdot \bar{D}$  is connected, and  $K$  is a continuum intersecting  $\beta$ , then  $M \cdot K$  is a continuum.*

*Proof.* By Theorem 3 of my paper, "Concerning a Continuum and Its Boundary,"<sup>2</sup>  $\beta$  is connected. Hence so is  $K + \beta$ . If  $M \cdot K$  is  $K$  or a subset of  $\beta$  then  $M \cdot K + \beta$  is  $K + \beta$  or  $\beta$ . If  $M \cdot K$  is neither  $K$  nor a subset of  $\beta$  then  $(K + \beta) - \beta$  is the sum of two mutually separated point sets  $M \cdot K - M \cdot K \cdot \beta$  and  $K - M \cdot K$ . Hence  $(M \cdot K - M \cdot K \cdot \beta) + \beta$  is connected. But this point set is identical with  $M \cdot K + \beta$ .

Theorem 1 does not remain true if, in the statement of its hypothesis, the stipulation that  $M$  is compact is replaced by the stipulation that  $K$  is compact even though there is added, to this hypothesis, the additional stipulation that the entire boundary of every component of  $M - \beta$  is a connected subset of  $\beta$  and the conclusion is so weakened as to require only that there exists a continuum containing  $M \cdot K$  and lying in  $M \cdot K + \beta$ . This may be seen with the help of Example 2 of C. B. However if, in the above proof of Theorem 1, Theorem 3 of C. B. is replaced by Theorem 4 of that paper, the resulting argument establishes the following result.

**THEOREM 2.** *Theorem 1 remains true if, in the statement of its hypothesis, the requirement that  $M$  be compact is replaced by the requirement that Axiom 2 hold true.*

**THEOREM 3.** *If, in a space satisfying Axioms 0 and 1, the continuum  $K$  contains  $\beta$ , the boundary of the compact continuum  $M$  and, for every component  $D$  of  $M - \beta$ ,  $K \cdot \bar{D}$  is a continuum, then  $M \cdot K$  is a continuum.*

**THEOREM 4.** *Theorem 3 remains true if the requirement that  $M$  be compact is replaced by the requirement that space satisfy Axiom 2.*

Theorems 3 and 4 may be easily established with the aid of Theorems 1 and 2, respectively, of C. B. Theorem 3 may also be easily proved with the help of Theorem 9 of my paper, "Concerning Accessibility."<sup>3</sup>

It may be seen from Example 2 of C. B. that Theorem 3 does not remain true if  $K$  instead of  $M$  is required to be compact.

**THEOREM 5.** *If, in a space satisfying Axioms 0 and 1,  $M$  is a continuum,  $K$  is a compact continuum intersecting  $M$  and, for every component  $E$  of  $S - M$  that intersects  $K$ ,  $T_E$  is a connected point set intersecting  $M \cdot K$  such that, for every component  $Q$  of  $K \cdot E$ , whose closure intersects  $M$ ,  $T_E$  intersects every component of  $\bar{Q} \cdot M$  and  $T$  is the sum of the point sets  $T_E$  for all  $E$ 's then  $M \cdot K + T$  is connected.*

*Proof.* Suppose  $M \cdot K + T$  is the sum of two mutually separated point sets  $F_1$  and  $F_2$ . Every component of  $T$  intersects  $M \cdot K$  and lies either in  $F_1$  or in  $F_2$ . Hence  $F_1 \cdot M \cdot K$  and  $F_2 \cdot M \cdot K$  exist. Since these point sets are mutually separated and their sum is the closed set  $M \cdot K$ , therefore they are closed. Hence  $K - M \cdot K$  contains a connected point set  $L$  such that  $\bar{L}$  intersects  $F_1$  and  $F_2$ . Let  $W$  denote the component of  $S - M$  which contains  $L$ . The point set  $T_W$  is a connected subset of  $T$  intersecting both  $F_1$  and  $F_2$ . This involves a contradiction.

*Theorem 5 does not remain true if the stipulation that  $K$  is compact is replaced by the stipulation that  $M$  is compact.* Consider the example obtained by interchanging  $M$  and  $K$  in Example 2 of C. B. But the following theorem holds.

**THEOREM 6.** *Theorem 5 remains true if the requirement that  $K$  be compact is replaced by the requirement that space satisfy Axiom 2.*

*Proof.* Suppose  $M \cdot K + T$  is the sum of the mutually separated point sets  $F_1$  and  $F_2$ . By a modification of Theorem 6 of my paper, "Concerning Domains Whose Boundaries Are Compact,"<sup>4</sup> there exists a component  $W$  of  $S - M$ , intersecting  $K$ , whose boundary intersects each of the closed point sets  $F_1 \cdot M \cdot K$  and  $F_2 \cdot M \cdot K$ . As in the above proof of Theorem 5, this leads to a contradiction.

The following theorem easily follows with the help of Theorem 6.

**THEOREM 7.** *If, in a space satisfying Axioms 0, 1 and 2, the continuum  $K$  contains the boundary of the continuum  $M$  and every component of  $S - M$  has a connected boundary then  $M \cdot K$  is connected.*

The following theorem may be easily established with the help of Theorem 5.

**THEOREM 8.** *If, in a space satisfying Axioms 0 and 1,  $M$  is a continuum and  $K$  is a compact continuum intersecting  $M$  and every component of  $S - M$  that intersects  $K$  has a connected boundary then  $M \cdot K$  plus the boundaries of all components of  $S - M$  that intersect  $K$  is connected.*

Theorem 8 does not remain true if its conclusion is replaced by " $M \cdot K$  is a subset of a component of  $M \cdot K$  plus the boundary of  $M$ ." Indeed it does not remain true if its hypothesis is strengthened by the addition of the stipulation that the boundary of every component of  $S - M$  is connected, and its conclusion is replaced by " $M \cdot K$  is a subset of some compact subcontinuum of  $M$ ." Consider the following example.

EXAMPLE 1. In a Cartesian plane  $E$  let  $T$  denote a totally disconnected compact and perfect point set lying on  $OY$  and such that the ordinate of its highest point is  $-10$ . Let  $A$  and  $B$  denote the points  $(-10, 0)$  and  $(10, 0)$ , respectively. Let  $H$  denote the set of all points of the graph of  $y = \sin(1/x)$  whose abscissae are neither less than  $-1/\pi$  nor greater than  $1/\pi$ . Let  $N$  denote the set of all points of  $OX$  that lie between  $A$  and  $(-1/\pi, 0)$  or between  $B$  and  $(1/\pi, 0)$ . Let  $M$  denote the point set  $O + H + N + A + B$  and let  $K$  denote the sum of all straight line intervals with one end-point at one of the points  $A$  and  $B$  and the other one at a point of  $T$ . Let  $S$  denote  $M + K$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . Here  $\beta$  consists of two points,  $A$  and  $B$ .

However, the following theorem holds true.

THEOREM 9. *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the continuum  $M$  and  $K$  is a compact continuum intersecting  $M$ , and the common part of  $M$  and the boundary of each component of  $S - M$  that intersects  $K$  is connected, then  $M \cdot K$  is a subset of a component of  $M \cdot K + \beta$ .*

THEOREM 10. *If, in a space satisfying Axioms 0 and 1,  $K$  is a compact continuum intersecting the continuum  $M$  and, for every component  $E$  of  $S - M$  that intersects  $K$ ,  $K \cdot \bar{E} \cdot M$  is a subset of some component of  $M \cdot K$  then  $M \cdot K$  is connected.*

*Proof.* For each component  $E$  of  $S - M$  that intersects  $K$ , let  $T_E$  denote the component of  $M \cdot K$  that contains  $K \cdot \bar{E} \cdot M$  and let  $T$  denote the sum of all  $T_E$ 's. By Theorem 5,  $M \cdot K + T$  is connected. But it is identical with  $M \cdot K$ .

It may be seen from the following example that Theorem 10 does not remain true if the requirement that  $K$  be compact is omitted even though the resulting hypothesis is strengthened by the addition of the stipulation that (1)  $K$  contains  $\beta$  and (2) every component of  $S - M$  is bounded by a connected subset of  $M$  and so is every component of  $M - \beta$  and, furthermore, if  $D$  is a component of  $M - \beta$ , every component of  $S - D$  is so bounded.

EXAMPLE 2. In a Cartesian plane  $E$ , let  $A$  and  $D$  denote the points  $(0, 2)$  and  $(0, 1)$  and, for each  $n$ , let  $A_n$  and  $B_n$  denote the points  $(1/n, 0)$  and  $(-1/n, 0)$ , respectively, and let  $AA_n$  and  $AB_n$  denote straight line intervals with end-points as indicated. Let  $M$  and  $K$  denote  $D + AB_1 + AB_2 + \dots$  and  $D + AA_1 + AA_2 + \dots$ , respectively. Let  $S$  denote  $M + K$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ .

Theorem 10 does not remain true if  $M$  instead of  $K$  is required to be compact. To see this, interchange  $M$  and  $K$  in the description of Example 2 of C. B.

Of Theorems 5-10, Theorem 7 is the only one that remains true if, in its statement,  $S - M$  is replaced by  $M - \beta$ .

The following theorem may be easily proved.



**THEOREM 11.** *If, in a space satisfying Axiom 0, the closed point set  $M$  intersects the connected point set  $K$  and  $M \cdot K$  is the sum of two mutually separated point sets  $H$  and  $L$  then both  $H$  and  $L$  intersect the boundary of  $M$ .*

**THEOREM 12.** *If, in a space satisfying Axioms 0 and 1, the compact point set  $\beta$  is the boundary of the compactly connected continuum  $M$  and, for every component  $D$  of  $M - \beta$  and component  $E$  of  $S - \bar{D}$ , the boundary of  $E$  is a subset of a component of  $\beta$  then if  $K$  is a compact continuum intersecting  $M$ ,  $M \cdot K$  is a subset of a component of  $M \cdot K + \beta$ .*

*Proof.* Let  $N$  denote the point set obtained by adding together all components of  $M \cdot K + \beta$  that intersect  $M \cdot K$ . Since  $M \cdot K$  and  $\beta$  are closed and compact,  $N$  is closed. Suppose it is the sum of two mutually exclusive closed point sets  $H$  and  $L$ . These point sets both intersect  $M \cdot K$ . Since the continuum  $K$  is compact it contains a connected point set  $T$  lying in  $S - M$  and such that  $\bar{T}$  intersects both  $H \cdot M \cdot K$  and  $L \cdot M \cdot K$ . Let  $A$  and  $B$  denote points belonging, respectively, to the subsets  $\bar{T} \cdot H \cdot M \cdot K$  and  $\bar{T} \cdot L \cdot M \cdot K$  of  $\beta$ . There exists a compact continuum  $M'$  lying in  $M$  and containing  $A$  and  $B$ . Suppose  $P$  is a point of  $M' - M' \cdot \beta$ . Let  $D_P$  denote the component of  $M - \beta$  that contains  $P$  and let  $E_P$  denote the component of  $S - \bar{D}_P$  that contains  $T$ . By hypothesis, the boundary of  $E_P$  is a subset of a component  $\beta_P$  of  $\beta$ . For each point  $P$  of  $M' - M' \cdot \beta$ ,  $\beta_P$  shields  $A + B$  from  $P$  in  $M$ . Hence, by Theorem 6 of C. B., there is a continuum containing  $A$  and  $B$  and lying in  $\beta$ . This involves a contradiction.

**THEOREM 13.** *Theorem 12 remains true if the requirement that  $K$  be compact is replaced by the requirement that  $M \cdot K$  be compact and that space satisfy Axiom 2.*

Theorem 13 may be proved by an argument identical with that given to prove Theorem 12 except for the substitution of "By Theorem 6 of D. C. B., there exists" for "Since the continuum  $K$  is compact it contains" in the third sentence of that argument.

Theorem 12 does not remain true on the omission of the stipulation that  $M$  is compactly connected even though it be stipulated that  $K$  contains  $\beta$ . Consider the following example.

**EXAMPLE 3.** In a Cartesian plane  $E$ , let  $O$  denote the origin and, for each positive integer  $n$ , let  $A_n, B_n, C_n, D_n$  and  $E_n$  denote the points  $(0, 1/n), (1/n, 1/n), (1/n, -1/n), (-1/n, -1/n)$  and  $(-1/n, 1)$  and let  $F_n$  denote a point lying midway between  $B_n$  and  $A_{n+1}$ . If  $X$  and  $Y$  are two points let  $XY$  denote the straight line interval whose extremities are  $X$  and  $Y$ . Let  $t_n$  denote the point set  $A_n B_n + B_n C_n + C_n D_n + D_n E_n$ . Let  $M$  and  $K$  denote the point sets  $O + t_1 + t_2 + t_3 + \dots$  and  $O + A_1 F_1 + F_1 A_2 + A_2 F_2 + F_2 A_3 + \dots$ , respectively. Let  $S$  denote the point set  $M + K$  and let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . Here  $\beta$  is the point set  $O + A_1 + A_2 + \dots$  and  $K$  is a compact con-



tinuum containing  $\beta$ . Furthermore if  $D$  is a component of  $M - \beta$  then, for some  $n$ ,  $D$  is  $t_n - A_n$  and if  $E$  is a component of  $S - \bar{D}_n$  then  $E$  is  $S - t_n$  and its boundary is the point  $A_n$  which is a connected subset of  $\beta \cdot \bar{D}$ . But  $M \cdot K = \beta$  and every component of  $M \cdot K + \beta$  is a point of the infinite point set  $\beta$ . Hence  $M \cdot K$  is not a subset of any component of  $M \cdot K + \beta$ .

Neither does Theorem 12 hold true if the stipulation that  $\beta$  is compact is replaced by the stipulation that Axiom 2 holds true. Consider the following example.

EXAMPLE 4. In a Cartesian space  $E$  of three dimensions, for each positive integer  $n$ , let  $T_n$  denote the portion of the graph of  $y = -n + 1/(x - x^2)$  that lies in the  $XY$  plane between the planes  $x = 0$  and  $x = 1$ . For each  $n$ , let  $S_n$  denote the set of all points  $(x, y, z)$  of this graph such that  $0 < x < 1$  and  $0 < z < 1/n$ . Let  $A$ ,  $B$  and  $C$  denote the points  $(0, 0, 2)$ ,  $(1, 0, 2)$  and  $(1, 0, 0)$ , respectively. Let  $K$  denote the arc obtained by adding together the straight line intervals  $OA$ ,  $AB$  and  $BC$ . Let  $M$  denote the  $XY$  plane and let  $S$  denote  $K + M + S_1 + S_2 + S_3 + \dots$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . Here  $K$  is compact and  $M \cdot K = O + C$ . The boundary of  $M$  is the point set obtained by adding together the open curves  $T_1, T_2, T_3, \dots$ , the  $Y$ -axis and a line through  $C$  parallel to the  $Y$ -axis. The point set  $M \cdot K + \beta$  is identical with  $\beta$  and no component of  $\beta$  contains both  $O$  and  $C$ . It is to be noted that  $\beta$  is not the sum of two mutually separated point sets containing  $O$  and  $C$ , respectively.

However, the following theorem holds true.

THEOREM 14. *If, in a space satisfying Axioms 0, 1, and 2, the continuum  $K$  contains  $\beta$ , the boundary of the continuum  $M$ , and, for every component  $D$  of  $M - \beta$  and component  $E$  of  $S - \bar{D}$ , the boundary of  $E$  is connected then  $M \cdot K$  is connected.*

*Proof.* Suppose  $M \cdot K$  is the sum of two mutually exclusive closed point sets  $H$  and  $L$ . By Theorem 11,  $H \cdot \beta$  and  $L \cdot \beta$  exist. Furthermore, by hypothesis,  $\beta$  is a subset of their sum. Hence, by Theorem 5 of C. B., there exists a component  $D$  of  $M - \beta$  such that  $\bar{D}$  intersects both  $H$  and  $L$ . The common part of the continua  $\bar{D}$  and  $K$  is the sum of the two mutually exclusive closed point sets  $D \cdot K \cdot H$  and  $\bar{D} \cdot K \cdot L$ . Therefore, by Theorem 6 of D. B. C., there exists an arc  $AB$  from the point  $A$  of  $\bar{D} \cdot K \cdot H$  to the point  $B$  of  $\bar{D} \cdot K \cdot L$  and having no point in common with  $\bar{D}$  except its end-points  $A$  and  $B$ . Let  $E$  denote the component of  $S - \bar{D}$  which contains  $AB - (A + B)$ . The boundary of  $E$  is a connected subset of  $\beta$  containing the points  $A$  and  $B$ . This involves a contradiction.

THEOREM 15. *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the continuum  $M$ ,  $D$  is a component of  $M - \beta$ ,  $K$  is a compact continuum intersecting  $\bar{D}$  but not lying wholly in it and  $Q$  is a component of  $K \cdot (S - \bar{D})$  then every component of  $\bar{Q} \cdot \bar{D}$  contains a point of  $\beta$ .*

*Proof.* Suppose the component  $H$  of  $\bar{Q} \cdot \bar{D}$  contains no point of  $\beta$ . Let  $O$  denote some point of  $Q - H$ . Since  $\bar{Q}$  is a compact continuum there exists a domain  $W$  containing  $H$ , but no point of  $O + \beta$ , and such that its boundary  $\gamma$  contains no point of  $\beta + \bar{Q} \cdot \bar{D}$ . The point set  $\bar{Q} \cdot W$  contains a connected point set  $T$  such that  $\bar{T}$  intersects  $\gamma$  and contains a point  $P$  of  $\bar{D}$ . Since  $P$  does not belong to  $\beta$ , it belongs to  $D$ . But  $T + P$  is a connected point set containing no point of  $\beta$ . Hence  $T + P$  is a subset of  $D$ . Since  $T$  is a subset of  $Q$  this involves a contradiction.

**THEOREM 16.** *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the compact continuum  $M$  and  $D$  is a component of  $M - \beta$  and, for every component  $E$  of  $S - \bar{D}$ , the common part of  $\beta$  and the boundary of  $E$  is connected and  $K$  is a compact continuum intersecting  $\bar{D}$  then  $K \cdot \bar{D}$  is a subset of a component of  $K \cdot \bar{D} + \beta$ .*

*Proof.* For every component  $E$  of  $S - \bar{D}$  that intersects  $K$ , let  $T_E$  denote the common part of  $\beta$  and the boundary of  $E$ . With the help of Theorem 15 it may be seen that if  $Q$  is a component of  $K \cdot E$  every component of  $Q \cdot \bar{D}$  intersects  $\beta$  and therefore  $T_E$ . Hence, by Theorem 5, if  $T$  is the sum of the continua  $T_E$  for all  $E$ 's,  $K \cdot \bar{D} + T$  is connected. But  $T$  is a subset of  $\beta$ .

**THEOREM 17.** *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the compact continuum  $M$  and, for every component  $D$  of  $M - \beta$  and component  $E$  of  $S - \bar{D}$ , the common part of  $\beta$  and the boundary of  $E$  is connected and  $K$  is a compact continuum containing  $\beta$  then  $M \cdot K$  is a continuum.*

*Proof.* If  $D$  is a component of  $M - \beta$ , the continuum  $K$  intersects  $\bar{D}$  and therefore, by hypothesis and Theorem 16,  $K \cdot \bar{D}$  is a subset of a component  $T_D$  of  $K \cdot \bar{D} + \beta$ . Let  $T$  denote the sum of all the point sets  $T_D$  for all  $D$ 's. By Theorem 1 of C. B.,  $\beta + T$  is connected. But this point set is identical with  $M \cdot K$ .

Theorem 17 remains true if the stipulation that the common part of  $\beta$  and the boundary of  $E$  is connected is replaced by the stipulation that it is a subset of a component of  $M \cdot K$ . But it does not remain true if "the common part of  $\beta$  and the boundary of  $E$ " is replaced by "the common part of  $\bar{D}$  and the boundary of  $E$ " or by "the boundary of  $E$ ." Consider the following example.

**EXAMPLE 5.** In a Euclidean plane  $E$  let  $\alpha$  denote a definite square and let  $\gamma$  denote a definite square enclosed by  $\alpha$ . Let  $O$  denote the mid-point of one side of  $\gamma$  and let  $Q$  denote a totally disconnected perfect point set lying on the opposite side of  $\gamma$ . Let  $M$  denote the point set obtained by adding together all straight line intervals with one end-point at  $O$  and the other one at a point of  $Q$ . Let  $K$  denote the sum of  $\alpha$  and  $\gamma$  and the set of all points that lie between them. Let  $S$  denote  $M + K$  and let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . Here  $S$  is compact,  $\beta$  is  $O + Q$ ,  $K$  contains  $\beta$  and, for every component  $D$  of  $M - \beta$ ,

$S - \bar{D}$  is connected and its boundary is  $\bar{D}$ . But  $M \cdot K$  is the uncountable totally disconnected point set  $\beta$ .

Theorem 17 does not remain true if the requirement that  $K$  be compact is omitted. To see this, interchange  $M$  and  $K$  in the description of Example 2 of C. B. It is clear from Example 3 that it does not remain true on the omission of the requirement that  $M$  be compact.

<sup>1</sup> *Amer. Math. Soc. Colloquium Pub.*, 13, New York (1932). The letter  $S$  denotes the set of all points.

<sup>2</sup> These PROCEEDINGS, 28, 550-555 (1942). This paper will be referred to as "C. B."

<sup>3</sup> These PROCEEDINGS, 25, 56-58 (1942).

<sup>4</sup> These PROCEEDINGS, 28, 555-581 (1942). This paper will be referred to as "D. C. B." If the proof of Theorem 6 of this paper is modified only to the extent of replacing the definition of  $H_n$  by the requirement that (1) for each  $n$  greater than 1,  $H_n$  be defined as it is there defined for each  $n$  and (2)  $H_1$  be a collection of connected domains properly covering  $K$  such that each of them lies in some region of  $G_1$  and every one of them that intersects  $M$  intersects  $M \cdot K$  and no one of them that intersects  $H$  has a point in common with any one of them that intersects  $L$ , then the arc  $C$  necessarily lies, except for its end-points, wholly in some component of  $S \cdot M$  that contains points of  $K$ .

## CONCERNING A CONTINUUM AND ITS BOUNDARY

BY R. L. MOORE

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF TEXAS

Communicated October 23, 1942

In this paper some theorems will be established concerning certain relationships between the boundary of a continuum and the components of that continuum minus its boundary. Numbered axioms and chapters herein referred to are axioms and chapters of the author's book "Foundations of Point Set Theory."<sup>1</sup>

**THEOREM 1.** *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the compact continuum  $M$  and, for every component  $D$  of  $M - \beta$ ,  $T_D$  is a connected point set lying in  $M$  and containing the common part of  $\beta$  and the boundary of  $D$ , and  $T$  is the sum of all the point sets  $T_D$  for all such  $D$ 's, then  $\beta + T$  is connected.*

*Proof.* Suppose  $\beta + T$  is the sum of two mutually separated point sets  $H$  and  $K$ . The point sets  $H \cdot \beta$  and  $K \cdot \beta$  exist and are mutually exclusive and closed. There exists a subcontinuum  $N$  of  $M$  which is irreducible from  $H \cdot \beta$  to  $K \cdot \beta$ . The connected point set  $N - (H \cdot \beta + K \cdot \beta)$  is a subset of some component  $L$  of  $M - \beta$ . Each of the point sets  $H \cdot \beta$  and  $K \cdot \beta$  con-

tains a limit point of  $N = (H \cdot \beta + K \cdot \beta)$  and therefore of  $L$ . Hence  $T_L$  intersects both  $H \cdot \beta$  and  $K \cdot \beta$ . This involves a contradiction.

**THEOREM 2.** *Theorem 1 remains true if, in the statement of its hypothesis, the requirement that  $M$  be compact is replaced by the requirement that space satisfy Axiom 2.*

*Proof.* Suppose  $\beta + T$  is the sum of two mutually separated point sets  $H_1$  and  $H_2$ . Suppose  $D$  is a component of  $M - \beta$ . Since the connected point set  $T_D$  is a subset of  $H_1 + H_2$ , it is a subset either of  $H_1$  or of  $H_2$ . If there exists at least one component  $D$  of  $M - \beta$  such that  $T_D$  is a subset of  $H_1$ , let  $H_1'$  denote  $H_1$  plus the sum of all such point sets  $D$ . Otherwise let  $H_1'$  denote  $H_1$ . Since Axiom 2 holds true the components of  $M - \beta$  are all domains and, by Theorem 2(b) of Chapter II, every limit point of their sum belongs either to one of them or to the closure of the sum of their boundaries. But, since they are domains, no one of them contains a limit point of its complement, and the boundary of every one of them is a subset of  $\beta$ . It follows that if the point  $P$  of  $H_2'$  is a limit point of  $H_1'$  then  $P$  belongs to  $H_2$  and is a limit point of  $H_1$ . This is impossible. Similarly  $H_1'$  contains no limit point of  $H_2'$ . Thus  $H_1'$  and  $H_2'$  are mutually separated. But their sum is the continuum  $M$ . This involves a contradiction.

**THEOREM 3.** *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the compact continuum  $M$  and, for every component  $D$  of  $M - \beta$ , the common part of  $\beta$  and the boundary of  $D$  is connected, then  $\beta$  is connected.*

**THEOREM 4.** *If, in a space satisfying Axioms 0, 1 and 2,  $\beta$  is the boundary of the continuum  $M$  and, for every component  $D$  of  $M - \beta$ , the common part of  $\beta$  and the boundary of  $D$  is connected, then  $\beta$  is connected.*

Theorems 3 and 4 are corollaries of Theorems 1 and 2, respectively.

If Axioms 0, 1 and 2 hold true and  $D$  is a component of the continuum  $M$  minus its boundary  $\beta$  then  $D$  is a domain and its boundary is a subset of  $\beta$ . Hence *Theorem 4 remains true if, in the statement of its hypothesis, "the common part of  $\beta$  and the boundary of  $D$ " is replaced by "the boundary of  $D$ ."* But this is not true of Theorem 3 even though the resulting hypothesis is strengthened by the addition of the stipulation that space is compact. Consider the following example.

**Example 1.** In a Cartesian plane  $E$  let  $Q$  denote a compact totally disconnected closed point set lying on the  $Y$ -axis, let  $A$  and  $B$  denote the points  $(-1, 0)$  and  $(-2, 0)$ , respectively, let  $S$  denote the sum of the straight line interval  $AB$  and all straight line intervals with one end-point at  $A$  and mid-point at a point of  $Q$ , let  $M$  denote the sum of all straight-line intervals with one end-point at  $A$  and the other one at a point of  $Q$  and let  $\beta$  denote the point set  $A + Q$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . In the compact space  $\Sigma$ , the boundary of the continuum  $M$  is  $\beta$  and if  $D$  is a component of  $M - \beta$  then the boundary of  $D$  is the continuum  $\bar{D}$ . But  $\beta$  is not connected.

That Theorem 3 does not remain true if the stipulation that  $M$  is compact is replaced by the stipulation that  $\beta$  is compact may be seen from the following example.

*Example 2.* In a Cartesian plane  $E$ , let  $O$  and  $A$  denote the points  $(0, 0)$  and  $(0, 2)$ , respectively, and, for each  $n$ , let  $B_n$  denote  $(-1/n, 0)$ . Let  $M$  denote the sum of  $O$  and the straight-line intervals  $AB_1, AB_2, AB_3, \dots$ . Let  $K$  denote a semi-circle with extremities at  $A$  and  $O$  and containing the point  $(1, 1)$ . Let  $S$  denote  $M + K$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . Here  $\beta$  is  $A + O$  and, for each component  $D$  of  $M - \beta$ ,  $\beta \cdot \bar{D}$  is  $A$ .

**THEOREM 5.** *If, in a space satisfying Axioms 0, 1 and 2,  $\beta$ , the boundary of the continuum  $M$  is the sum of two mutually exclusive closed point sets  $\beta_1$  and  $\beta_2$  then there is a component of  $M - \beta$  whose closure intersects both  $\beta_1$  and  $\beta_2$ .*

*Proof.* The closure of every component of  $M - \beta$  intersects  $\beta$  and therefore either  $\beta_1$  or  $\beta_2$ . Let  $M_1$  denote the set of all points that belong either to  $\beta_1$  or to some component of  $M - \beta$  whose closure intersects  $\beta_1$ . Since their sum is the connected point set  $M$ ,  $M_1$  and  $M_2$  are not mutually separated. Suppose they are mutually exclusive. Then one of them contains a limit point of the other one. But every component of  $M - \beta$  is a domain and no domain contains a limit point of its complement. Hence either  $\beta_1$  contains a limit point of  $M_2$  or  $\beta_2$  contains a limit point of  $M_1$ . Suppose  $\beta_1$  contains a point  $P$  which is a limit point of  $M_2$ . Let  $Q$  denote the collection of all components of  $M - \beta$  whose closures intersect  $\beta_2$ . Since it is not a limit point of  $\beta_2$ ,  $P$  must be a limit point of  $Q^*$ . Hence, by Theorem 2(b) of Chapter II,  $P$  belongs to the closure of the sum of the boundaries of the domains of the collection  $Q$ . Since the boundary of every domain of  $Q$  belongs to the closed point set  $\beta_2$  it follows that  $P$  belongs to  $\beta_2$  and therefore that  $\beta_1$  intersects  $\beta_2$ . The same result is obtained in case  $\beta_2$  contains a limit point of  $\beta_1$ . Thus the supposition that  $M_1$  and  $M_2$  are mutually exclusive leads to a contradiction. It follows that there is a component of  $M - \beta$  whose closure intersects  $\beta_1$  and  $\beta_2$ .

*Definition.* If  $H, L$  and  $N$  are subsets of the continuum  $M$ ,  $N$  is said to *shield*  $H$  from  $L$  in  $M$  if  $N$  has no point in common with  $L$  but intersects every connected subset of  $M$  which intersects both  $H$  and  $L$ .

**THEOREM 6.** *If  $A$  and  $B$  are points of the continuum  $M$ ,  $T$  is a compact subcontinuum of  $M$  containing  $A$  and  $B$ ,  $H$  is a closed and compact subset of  $M$  and, for each point  $P$  of  $T - (A + B + T \cdot H)$  there is a continuum lying in  $H$  and shielding  $A + B$  from  $P$  in  $M$  then  $A$  and  $B$  belong to the same component of  $H$ .*

*Proof.* Suppose  $A$  and  $B$  do not lie in the same component of  $H$ . There exists a well-ordered sequence  $\alpha$  whose terms are the points of  $T - (A + B + T \cdot H)$ , a well-ordered sequence  $\beta_1$  whose terms are the continua which

are subsets of  $H$ , and a well-ordered sequence  $\beta_2$  whose terms are the continua that lie in  $T + H$ . Suppose  $T$  is not a subset of  $H$ . There exists a well-ordered sequence  $\gamma$  such that (1)  $T$  is the first term of  $\gamma$ , (2) each term of  $\gamma$  is a continuum lying in  $T + H$  but not wholly in  $H$  and (3) if  $x$  is a term of  $\gamma$  and  $P_x$  is the first point of  $\alpha$  belonging to  $x$  and  $H_x$  is the first term of  $\beta_1$  which shields  $A + B$  from  $P_x$  in  $M$ ,  $T_{Ax}$  is the first term of  $\beta_2$  which is an irreducible continuum from  $A$  to  $H_x$  lying in  $x$  and  $T_{Bx}$  is the first term of  $\beta_2$  which is an irreducible continuum from  $B$  to  $H_x$  lying in  $x$ , then  $x$  is the last term of  $\gamma$  or  $H_x + T_{Ax} + T_{Bx}$  is the first term following  $x$  in  $\gamma$ , according as  $T_{Ax} + T_{Bx}$  is, or is not, a subset of  $H$ , (4) if  $\gamma'$  is a well-ordered proper subsequence of  $\gamma$  with no last term and  $\gamma'$  has, as one of its terms, each term of  $\gamma$  which, in  $\gamma$ , precedes a term of  $\gamma'$ , then the limiting set of the sequence  $\gamma'$  is the first term of  $\gamma$  which, in  $\gamma$ , follows all the terms of  $\gamma'$ , unless this limiting set is a subset of  $H$  in which case  $\gamma'$  is  $\gamma$ .

Let  $\gamma_T$  denote a well-ordered sequence such that (1)  $x$  is a term of  $\gamma_T$  if, and only if,  $x$  is the common part of  $T - T \cdot H$  and some continuum which is a term of  $\gamma$ , (2)  $x$  precedes  $y$  in  $\gamma_T$  if, and only if, there exist terms  $x'$  and  $y'$  of  $\gamma$  such that  $x = x' \cdot T$  and  $y = y' \cdot T$  and  $x'$  precedes  $y'$  in  $\gamma$ . Let  $\gamma_H$  denote a well-ordered sequence described in exactly the same manner except for the substitution of  $\gamma_H$  for  $\gamma_T$  and of  $H$  for  $T - T \cdot H$ . Let  $Q$ ,  $Q_T$  and  $Q_H$  denote the limiting sets of  $\gamma$ ,  $\gamma_T$  and  $\gamma_H$ , respectively. It may be seen that  $Q = Q_T + Q_H$ . Suppose  $Q_T$  is not a subset of  $H$ . Every term of  $\gamma_T$  contains every one that follows it and therefore  $Q_T - H \cdot Q_T$  is the common part of all the point sets of this sequence. Let  $P$  denote the first point of  $Q_T$  in the sequence  $\alpha$ . Let  $\omega$  denote a subsequence of  $\gamma$  such that  $x$  belongs to  $\omega$  if, and only if, it contains a point that precedes  $P$  in  $\alpha$ . The second term of  $\gamma$  that follows every term of  $\omega$  exists but does not contain  $P$ . This involves a contradiction.

Theorem 6 does not remain true if the stipulation that, for each point  $P$  of  $T - (A + B + T \cdot H)$  there is a continuum lying in  $H$  and shielding  $A + B$  from  $P$  is replaced by the stipulation that for each such point  $P$  there is a continuum lying in  $H$  and intersecting every subcontinuum of  $M$  which intersects both  $P$  and  $A + B$ . Consider the following example.

*Example 3.* In a Euclidean plane, let  $M$  denote a compact indecomposable continuum and let  $A$  and  $B$  denote points belonging to different composants of  $M$ . Let  $H$  denote  $A + B$ . If  $P$  is a point of  $M - H$ ,  $M$  is an irreducible continuum either from  $P$  to  $A$  or from  $P$  to  $B$ . In the first case,  $B$  intersects every subcontinuum of  $M$  that intersects both  $P$  and  $A + B$ . In the second case,  $A$  intersects every such continuum. But no component of  $H$  contains both  $A$  and  $B$ .

The following theorem may be easily proved with the assistance of Theorem 6.



**THEOREM 7.** *If the points  $A$  and  $B$  belong to different components of  $\beta$ , the boundary of the compact continuum  $M$ , there exists a component  $D$  of  $M - \beta$  such that no component of  $\beta$  shields  $A + B$  from  $D$  in  $M$ .*

**THEOREM 8.** *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the continuum  $M$ ,  $D$  is a component of  $M - \beta$  and  $K$  is a continuum intersecting  $D$  but not lying wholly in it and  $M \cdot K$  is compact then every component of  $K \cdot \bar{D}$  contains a point of  $\beta$ .*

*Proof.* Suppose there exists a component  $N$  of  $K \cdot \bar{D}$  containing no point of  $\beta$ . Then there exists a subset  $H$  of  $K \cdot \bar{D} \cdot (S - \beta)$  containing  $N$  such that either  $N = H$  or  $H$  and  $K \cdot \bar{D} - H$  are mutually exclusive and closed. There exists a domain  $I$  containing  $H$  such that  $I \cdot K \cdot \bar{D} = H$  and  $I \cdot (\beta + S - M)$  is vacuous. There exists a subcontinuum  $T$  of  $K$  lying in  $I$  and intersecting both  $N$  and  $S - I$ . The point set  $T$  is a connected subset of  $K$  intersecting  $N$  but not lying wholly in it. Hence it is not a subset of  $D$ . But it is a subset of  $M - \beta$  and it contains a point of  $D$ . This involves a contradiction.

Theorem 8 does not remain true if the stipulation that  $M \cdot K$  is compact is replaced by the stipulation that  $\bar{D}$  is compact and that  $M$  is identical with  $K$ . Consider the following example.

**Example 4.** In a Cartesian plane  $E$ , let  $A$  and  $D$  denote the points  $(0, 2)$  and  $(0, 1)$  and, for each  $n$ , let  $B_n$  denote the point  $(-1/n, 0)$  and let  $AB_n$  denote the straight-line interval whose end-points are  $A$  and  $B_n$ . Let both  $K$  and  $M$  denote the point set  $D + AB_2 + AB_3 + \dots$  and let  $S$  denote  $M + AB_1$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . Here  $\beta$  is  $A$  and the point  $D$  is a component of  $M - \beta$ . Though  $\bar{D}$  is compact, the component  $D$  of  $K \cdot \bar{D}$  does not contain  $A$ .

That Theorem 8 does not remain true if the requirement that  $M \cdot K$  be compact is replaced by the requirement that space be the plane and  $D$  be the only component of  $M - \beta$  may be seen from the following example.

**Example 5.** In a Cartesian plane  $E$ , let  $M$  denote the set of all points  $(x, y)$  such that either  $-1/x^2 \leq y \leq 1/x^2$  or  $x = 0$  and let  $K$  denote the set of all points  $(x, y)$  such that either  $x = 0$  or  $y = (2 \sin 1/x^2)/x^2$ .

However, Theorem 8 does remain true if the requirement that  $M \cdot K$  be compact is replaced by the requirement that Axiom 2 hold true and  $K \cdot \bar{D}$  be compact.

**THEOREM 9.** *If, in a space satisfying Axioms 0 and 1,  $\beta$  is the boundary of the continuum  $M$ ,  $D$  is a component of  $M - \beta$  and  $K$  is a continuum intersecting  $\bar{D}$  and such that  $M \cdot K$  is compact and either  $\beta \cdot \bar{D}$  is vacuous or there exists a subcontinuum of  $K$  lying in  $\bar{D}$  and containing  $\beta \cdot \bar{D}$ , then  $K \cdot \bar{D}$  is a continuum.*

Theorem 9 may be easily established with the help of Theorem 8. It does not remain true if  $M \cdot K$  is replaced by  $\bar{D}$ . Consider the following example.

*Example 6.* Let space be that of Example 2, but now let  $K$  denote the point set  $AB_1 + AB_2 + \dots$ , let  $M$  denote  $A + (S - AB_1)$  and let  $D$  denote  $S - (AB_1 + AB_2) + \dots$ . The point set  $M \cdot K$  is now  $K$ ,  $\bar{D}$  is compact and  $\beta$ , the boundary of  $M$ , is the point  $A$  which is a subcontinuum of  $K$  lying in  $\bar{D}$  and coinciding with  $\beta \cdot \bar{D}$ . But  $K \cdot \bar{D}$  is  $A + O$  which is not connected.

Theorem 9 does not remain true if the stipulation that  $M \cdot K$  is compact is replaced by the stipulation that  $\bar{D}$  is compact even though the resulting hypothesis is strengthened by the elimination of the word "either" and "or there exists a subcontinuum of  $K$  lying in  $\bar{D}$  and containing  $\beta \cdot \bar{D}$ ." Consider the following example.

*Example 7.* In a Cartesian plane  $E$ , let  $A$ ,  $A_1$  and  $A_2$  denote the points  $(0, 4)$ ,  $(0, 1)$  and  $(0, 3)$ , respectively and, for each  $n$ , let  $B_n$  denote  $(-1/n, 0)$  and let  $AB_n$  denote the straight-line interval whose extremities are  $A$  and  $B_n$ . Let  $T$  denote a semi-circle passing through  $(1, 2)$  and with its extremities at  $A_1$  and  $A_2$ . Let  $K$  denote the point set  $A_1 + A_2 + AB_2 + AB_3 + \dots$ , let  $M$  denote  $K + T$  and let  $S$  denote  $M + AB_1$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . Here  $D$  is  $T - (A_1 + A_2)$ ,  $\beta$  is  $A$  and  $\beta \cdot \bar{D}$  is vacuous. But  $K \cdot \bar{D}$  is  $A_1 + A_2$  which is not connected.

**THEOREM 10.** *Theorem 9 remains true if the stipulation that  $M \cdot K$  is compact is replaced by the stipulation that Axiom 2 holds true.*

*Proof.* Since Axiom 2 holds true,  $D$  is a domain whose boundary is  $\beta \cdot \bar{D}$ . Hence, by hypothesis, there exists a subcontinuum  $H$  of  $K$  lying in  $\bar{D}$  and containing  $\beta \cdot \bar{D}$ . If  $K \cdot \bar{D}$  is neither  $H$  nor  $K$  then  $K - H$  is the sum of the point sets  $(K - H) \cdot D$  and  $K - K \cdot \bar{D}$ . Since  $D$  is a domain these point sets are mutually separated. Hence  $(K - H) \cdot D + H$  is a continuum. But this point set is  $K \cdot \bar{D}$ .

<sup>1</sup> *Amer. Math. Soc. Colloquium Pub.*, 13, New York (1938).

## CONCERNING DOMAINS WHOSE BOUNDARIES ARE COMPACT

BY R. L. MOORE

DEPARTMENT OF PURE MATHEMATICS, UNIVERSITY OF TEXAS

Communicated October 23, 1942

The following theorem may be established by an argument differing slightly from that given to prove Theorem 12 of Chapter IV of the author's "Foundations of Point Set Theory."<sup>1</sup>

**THEOREM 1.** *If, in a space satisfying Axioms 0-5,  $H$  and  $K$  are two mutually exclusive closed and compact point sets and  $L$  is a closed point set*



containing no point of  $H$  and  $G$  is a collection of regions covering  $H$  there exists a domain  $D$  containing  $H$  and such that (1)  $\bar{D}$  contains no point of  $K$  and the boundary of  $D$  is a compact point set containing no point of  $L$ , (2)  $D$  is the sum of a finite number of mutually exclusive connected domains  $\bar{Q}$  such that (a) every complementary domain of  $\bar{Q}$  is bounded by a simple closed curve, (b)  $\bar{Q}$  is the sum of a finite number of domains each bounded by a simple closed curve lying in some region of  $G$ .

**THEOREM 2.** *If Axioms 0-3 hold true,  $\omega$  is a point and  $M$  is a closed and compact point set not containing  $\omega$  and  $W_1, W_2, W_3, \dots$  is a sequence satisfying with respect to  $M$  all the requirements of Theorem 81 of Chapter I then, for each closed and compact point set  $T$  which does not intersect  $M$ , there exists a number  $\delta$  such that if a closed point set  $K$  separates a point  $P$  of  $M$  from  $\omega$  and lies in some region of  $W_\delta$  which contains  $P$  then  $K$  separates  $P$  from  $T$ .*

*Proof.* Suppose there exists a closed and compact point set  $T$  for which there is no such number  $\delta$ . Then, for each positive integer  $n$ , there exists a region  $g_n$  belonging to  $W_n$  and containing both a Point  $P_n$  of  $M$  and a closed point set which separates  $P_n$  from  $\omega$  but not from  $T$ . There exists an ascending sequence of positive integers  $n_1, n_2, n_3, \dots$  such that if, for each  $i$ ,  $P_i'$  denotes the term of the sequence  $P_1, P_2, P_3, \dots$  whose subscript is  $n_i$ , then  $P_1', P_2', P_3', \dots$  converges to a point  $P$  of  $M$ . For each  $i$ , let  $W_i'$  denote the term of the sequence  $W_1, W_2, W_3, \dots$  whose subscript is  $n_i$  and let  $g_i'$  denote the term of  $g_1, g_2, g_3, \dots$  with the same subscript. By Axiom 3, the domain  $S - P$  is connected. Hence, by a theorem of Dr. Harlan Cross Miller's,<sup>2</sup> there exists a continuum  $L$  containing  $T + \omega$  and lying in  $S - P$ . For each  $n$ , there exists a closed point set  $K_n$ , lying in  $G_n'$ , such that  $S - K_n$  is the sum of two mutually separated point sets one containing  $\omega$  and the other one containing both  $P_n'$  and some point of  $T$ , and therefore of  $L$ . Since  $L$  is connected it contains a point  $X_n$  belonging to  $K_n$  and therefore to  $g_n'$ . Since, for each  $n$ ,  $X_n$  belongs to a region of  $W_n'$ , and therefore to a region of  $W_n$ , some subsequence  $\alpha$  of the sequence  $X_1, X_2, X_3, \dots$  converges to a point  $X$  of  $M$ . Since each point of  $\alpha$  belongs to  $L$  therefore  $X$  belongs also to  $L$ . Hence it is distinct from  $P$ . Therefore there exists a number  $m$  such that if  $x, y$  and  $z$  are intersecting regions belonging to one of the collections  $W_m, W_{m+1}, W_{m+2}, \dots$  and  $x$  intersects  $y$  and  $y$  intersects  $z$  and  $x$  contains  $P$  then  $z$  does not contain  $X$ . There exist regions  $x$  and  $z$  belonging to  $W_m$  and containing  $P$  and  $X$ , respectively. There exists a number  $j$ , greater than  $m$ , such that  $P_j'$  belongs to  $x$  and  $X_j$  belongs to  $z$ . The region  $g_j'$  contains  $P_j'$  and  $X_j$  and lies in some region  $y$  of  $W_m$ . Since  $x$  intersects  $y$  and  $y$  intersects  $z$ , the region  $z$  does not contain  $X$ . This involves a contradiction.

The following theorem may be established with the help of Theorems 1 and 2 and Theorem 81 of Chapter I.

**THEOREM 3.** *If, in a space satisfying Axioms 0-5,  $M$  is a closed and com-*

compact point set there exists a sequence of domains  $D_1, D_2, D_3, \dots$  such that (1) for each  $n$ ,  $D_n$  contains  $M$ , and  $\bar{D}_{n+1}$  is a subset of  $D_n$  and the boundary of  $D_n$  is compact, (2) for each  $n$ ,  $D_n$  is the sum of a finite number of mutually exclusive connected domains  $Q$  such that every complementary domain of  $\bar{Q}$  is bounded by a simple closed curve, (3) if  $D$  is a domain containing  $M$  there exists a number  $\delta$  such that, for every  $n$  greater than  $\delta$ , the boundary of  $D_n$  is a subset of  $D$ , (4) if  $D$  is a domain containing  $M$  and bounded by a compact point set then, for some  $n$ ,  $D_n$  is a subset of  $D$ , (5)  $M$  is the common part of the domains of this sequence.<sup>3</sup>

AXIOM  $\bar{5}$  (F. B. Jones).<sup>4</sup> If  $A$  is a point of a region  $R$  and  $B$  is a point distinct from  $A$  there exists in  $R$  a closed and compact point set  $T$  separating  $A$  from  $B$ .

THEOREM 4. If, in a space satisfying Axioms 0-3 and  $\bar{5}$ ,  $M$  is a closed and compact point set there exists a sequence of domains  $D_1, D_2, D_3, \dots$  such that (1) for each  $n$ ,  $\bar{D}_{n+1}$  is a subset of  $D_n$ , (2) for each  $n$ , the boundary of  $D_n$  is compact, (3) if  $D$  is a domain containing  $M$  there exists a number  $\delta$  such that, for every  $n$  greater than  $\delta$ , the boundary of  $D_n$  is a subset of  $D$ , (4) if  $D$  is a domain containing  $M$  and bounded by a compact point set then, for some  $n$ ,  $D_n$  is a subset of  $D$ , (5)  $M$  is the common part of the domains of this sequence.

Theorem 4 may be established with the aid of Theorem 81 of Chapter I, Theorem 2, and a theorem corresponding to Theorem 1.

THEOREM 5. In a space satisfying Axioms 0-3 and  $\bar{5}$ , no domain whose boundary is compact has uncountably many components.

*Proof.* Suppose there exists a domain  $D$  bounded by a compact point set  $M$  and having uncountably many components. There exists a sequence of domains  $D_1, D_2, D_3, \dots$  satisfying, with respect to  $M$ , all the conditions of Theorem 4. If  $x$  is a component of  $D$  there exists a positive integer  $n_x$  such that  $\bar{D}_{n_x}$  does not contain  $x$ . There exists a positive integer  $k$  such that, for uncountably many components  $x$  of  $D$ ,  $n_x$  is  $k$ . For each such component  $x$  of  $D$ , the common part of  $x$  and  $\beta_k$ , the boundary of  $D_k$ , exists and is a domain with respect to  $\beta_k$ . But since  $\beta_k$  is compact there do not exist uncountably many mutually exclusive domains with respect to it. This involves a contradiction.

It is not true that if Axioms 0, 1, 2 and  $\bar{5}$  hold true then, for each point  $P$ , there exists a sequence of domains  $D_1, D_2, D_3, \dots$ , each having a compact boundary, such that  $P$  is their common part or that no domain with a compact boundary has uncountably many components. Consider the following example.

*Example 1.* Suppose  $E$  is a Cartesian space of three dimensions. Interpret the word point to mean a point of  $E$  and call a point set  $M$  a region if and only if either  $M$  is the interior of a sphere with center at  $O$  or  $M$  is a segment, not containing  $O$ , of some straight line through  $O$ . There exists no countable collection of domains with compact boundaries such that  $O$

is their common part. Furthermore the set of all points distinct from  $O$  is a domain whose boundary is  $O$  and this domain has uncountably many components.

**THEOREM 6.** *If, in a space satisfying Axioms 0, 1 and 2, the common part of the continuum  $K$  and the closed point set  $M$  is the sum of the two mutually exclusive closed point sets  $H$  and  $L$  then there exists an arc having in common with  $M$  only its end-points which belong to  $H$  and  $L$ , respectively.*

*Proof.* For each  $n$ , let  $H_n$  denote the collection of all connected domains  $D$  such that (1)  $D$  is a subset of some region of  $G_n$  and (2)  $D$  either intersects  $M \cdot K$  or does not intersect  $M$ . There exists a simple chain  $C_1$  of domains of  $H_1$  such that (1)  $M$  intersects no link of  $C_1$  except the first one and the last one, (2) the first link of  $C_1$  intersects  $H$  and the last one intersects  $L$ . There exists<sup>6</sup> a simple chain  $C_2$  of domains of  $H_2$  such that (1) the closure of each link of  $C_2$  is a subset of some link of  $C_1$ , (2) if a link  $x$  of  $C_2$  lies in a link  $y$  of  $C_1$  then every link that follows  $x$  in  $C_2$  lies either in  $y$  or in some link that follows  $y$  in  $C_1$ , (3)  $M$  intersects no link of  $C_2$  except the first one and the last one, and (4) the first one and last one intersect  $H$  and  $L$ , respectively. This process may be continued. Thus there exists an infinite sequence of simple chains  $C_1, C_2, C_3, \dots$  such that (1) each link of  $C_n$  belongs to  $H_n$ , (2) the closure of each link of  $C_{n+1}$  is a subset of some link of  $C_n$ , (3) if a link  $x$  of  $C_{n+1}$  lies in a link  $y$  of  $C_n$  then each link that follows  $x$  in  $C_{n+1}$  lies either in  $y$  or in some link that follows  $y$  in  $C_n$ , (4) the first link of  $C_n$  intersects  $H$  and the last one intersects  $L$  and no other one intersects  $M$ . For each  $n$ , let  $C_n^*$  denote the sum of all the links of  $C_n$ . Let  $C$  denote the common part of the point sets  $C_1^*, C_2^*, C_3^*, \dots$ . For each  $n$ , let  $\alpha_n$  and  $\beta_n$  denote the first and last links, respectively, of the chain  $C_n$ . For each  $n$ ,  $\alpha_n$  contains  $\bar{\alpha}_{n+1}$  and  $\alpha_n$  is a subset of some region of  $G_n$ . Hence the point sets  $\alpha_1, \alpha_2, \alpha_3, \dots$  have only one point in common. Call this point  $A$ . Let  $B$  denote the point common to the point sets  $\beta_1, \beta_2, \beta_3, \dots$ . The points  $A$  and  $B$  belong to  $H$  and  $L$ , respectively, and, for each  $n$ ,  $C_n$  is a simple chain from  $A$  to  $B$ . The point set  $C$  is<sup>6</sup> an arc from  $A$  to  $B$ . Clearly  $C \cdot M = A + B$ .

**THEOREM 7.** *If, in a space satisfying Axioms 0, 1 and 2,  $M$  is a continuum such that every component of  $S - M$  has a connected boundary and  $K$  is a continuum containing the boundary of  $M$  then  $M \cdot K$  is a continuum.*

*Proof.* Suppose  $M \cdot K$  is the sum of two mutually exclusive closed point sets  $H$  and  $L$ . By Theorem 6 there is a component  $D$  of  $S - M$  whose boundary  $\gamma$  intersects both  $H$  and  $L$ . The point sets  $\gamma \cdot H$  and  $\gamma \cdot L$  are subsets of  $\gamma$ , and  $\gamma$  is a connected subset of  $M \cdot K$ . Thus there is a component of  $M \cdot K$  intersecting both  $H$  and  $L$ . This involves a contradiction.

**THEOREM 8.** *If, in a space satisfying Axioms 0, 1 and 2,  $M$  is a closed point set with a compact boundary  $\beta$  then there exists a closed and compact point set  $M'$  lying in  $M$ , containing  $\beta$  and such that if  $D$  is a component of*

$M - \beta$  such that every component of  $S - \bar{D}$  has a connected boundary then  $M' \cdot \bar{D}$  is a continuum containing the boundary of  $D$ .

*Proof.* There exists<sup>3</sup> a compact continuum  $N$  containing  $\beta$ . Let  $M'$  denote  $M \cdot N$ . Suppose  $D$  is a component of  $M - \beta$  such that every component of  $S - \bar{D}$  has a connected boundary. The boundary of  $D$  is a subset of  $M'$  since it is a subset of  $\beta$ . By Theorem 7,  $M' \cdot D$  is a continuum.

**THEOREM 9.** *If, in a space satisfying Axioms 0, 1 and 2,  $O$  is a point of a continuum  $M$  whose boundary  $\beta$  is compact and every component of the complement of  $M$  has a connected boundary, but  $M$  is not locally connected at  $O$  then there exist a domain  $D$  containing  $O$  and a sequence  $M_1, M_2, M_3, \dots$  of mutually exclusive subcontinua of  $M$  such that (1) for each  $n$ ,  $M_n$  is a component of  $M \cdot \bar{D}$  containing a point of  $D$  and a point of the boundary of  $D$ , (2) the sequence  $M_1, M_2, M_3, \dots$  converges to a subcontinuum of  $M$  containing  $O$  and a point of the boundary of  $D$ .*

*Proof.* There exist a domain  $D$  containing  $O$  and a sequence of distinct points  $P_1, P_2, P_3, \dots$  belonging to  $M \cdot D$  such that (1) this sequence converges to  $O$  and (2) no component of  $M \cdot \bar{D}$  contains more than one of the points  $O, P_1, P_2, P_3, \dots$ . There exists a compact continuum  $T$  containing  $\beta + O + P_1 + P_2 + P_3 + \dots$ . Let  $L$  denote  $T \cdot M$ . By hypothesis and Theorem 7,  $L$  is connected. For each  $n$ ,  $L$  contains  $P_n$  and  $O$  but no connected subset of  $D$  containing  $P_n$  and  $O$ . Hence it contains a point of  $S - D$ . But  $L$  is compact. Hence  $L_n$ , the component of  $L \cdot \bar{D}$  that contains  $P_n$ , contains a point of the boundary of  $D$ . The point sets  $L_1, L_2, L_3, \dots$  are mutually exclusive subsets of the closed and compact point set  $L \cdot \bar{D}$ . It follows that there exists an increasing sequence of positive integers  $n_1, n_2, n_3, \dots$  such that  $L_{n_1}, L_{n_2}, L_{n_3}, \dots$  converges to a compact continuum containing  $O$  and lying in  $L \cdot \bar{D}$  and intersecting the boundary of  $D$ .

Theorem 9 does not remain true on the omission from its hypothesis of the condition requiring that Axiom 2 be satisfied. Consider the following example.

**Example 2.** In a Cartesian plane  $E$ , let  $O, A, B, C, D$  and  $E$  denote the points  $(0, 0), (0, -1), (1, -1), (2, -1), (2, 0)$  and  $(1, -2)$ , respectively, and let  $OA, AC, CD$  and  $BE$  denote straight line intervals with end-points as indicated. Let  $K$  denote the set of all points of the graph of  $y = (1/x) \sin(1/x)$  whose abscissae lie between 0 and 2, let  $M$  denote the point set  $K + OA + AC + CD$  and let  $S$  denote  $M + BE$ . Let  $\Sigma$  denote the subspace of  $E$  whose points are the points of  $S$ . In this space the continuum  $M$  is not locally connected at  $O$ . Its boundary consists of the single point  $B$  and the only component of  $S - M$  has  $B$  as its boundary. But  $M$  has no continuum of condensation.

Theorem 9 does not remain true on the omission of the requirement that every component of the complement of  $M$  have a connected boundary.

This may be seen with the aid of Example 1 of my paper "Concerning Accessibility."<sup>2</sup>

**THEOREM 10.<sup>7</sup>** *If, in a space satisfying Axioms 0, 1 and 2,  $M$  is a continuous curve and  $N$  is a subcontinuum of  $M$  whose boundary with respect to  $M$  has no continuum of condensation and every component of  $S - N$  has a connected boundary then  $N$  is a continuous curve.*

Theorem 10 may be established by an argument identical with that given in lines 3 to 27 of page 114, in Chapter II, except for the substitution of  $M'$  for  $N$  in line 24 and of "by Theorem 9 of the present paper" in place of "with the aid of Theorem 22, by a modification of the argument employed to prove Theorem 8."

**THEOREM 11.** *If, in a space satisfying Axioms 0-5,  $M$  is a continuum whose boundary is compact then  $M$  is compactly connected.*

*Proof.* The boundary of  $M$  is a subset of a compact continuum  $K$ . Since the boundary of every component of  $S - M$  is connected therefore, by Theorem 7,  $M \cdot K$  is a continuum. That  $M$  is compactly connected follows as in the proof of Theorem 10 of "Concerning Accessibility."

**THEOREM 12.** *If, in a space satisfying Axioms 0-5,  $M$  is a continuous curve and  $N$  is a subcontinuum of  $M$  whose boundary has no continuum of condensation then  $N$  is a continuous curve.*

Theorem 12 may be easily established with the aid of Theorem 10 and the fact that, in a space satisfying Axioms 0-5, every compact boundary of a complementary domain of a continuum is itself a continuum.

In the hypotheses of some of the numbered theorems of Chapter IV it is stipulated that certain domains form a semi-contracting set. With the aid of the results obtained in this paper it may be seen that all of these theorems remain true on the omission of this stipulation.

<sup>1</sup> *Am. Math. Soc. Colloquium Pub.*, Vol. XIII, New York (1932). Theorem 10 may be strengthened in a similar manner by the interpolation of "and  $L$  is a closed point set containing no point of  $M$ " between " $M$ " and "there" and the substitution, at the end, of " $L$  or of  $M$ " in place of " $M$ ." That Theorem 13 remains true on the removal of the requirement that  $K$  be compact may be shown by an argument identical with that given in Chapter IV (to prove this theorem in its original form) except for the substitution of the following two sentences, in place of the second sentence of that argument: "By Theorem 41 of Chapter I, there exists a domain  $Q$  containing  $H$  but no point of  $K + T$  and such that  $\bar{Q} - Q$  contains no point of  $M + K + T$ . Let  $L$  denote  $M \cdot Q$  and let  $N$  denote  $(M + K) - L$ . The point set  $M + K + T$  is the sum of the two point sets  $L$  and  $N + T$  which contain  $H$  and  $K + T$ , respectively." Except in the case of Axiom 5, numbered axioms and chapters herein referred to are axioms and chapters of this book. The letter  $S$  denotes the set of all points.

<sup>2</sup> Miller, H. C., "A Theorem Concerning Closed and Compact Point Sets Which Lie in Connected Domains," *Bull. Am. Math. Soc.*, 46, 520-521 (1940).

<sup>3</sup> It does not follow even for the case where  $M$  is a single point, that if  $D_n$  is a domain containing  $M$  then, for some  $n$ ,  $D_n$  is a subset of  $D$ . Cf. Example 2 of my paper "Concerning Accessibility," these PROCEEDINGS, 25, 646-653 (1939). In the space of this

example there does not exist a sequence of domains  $D_1, D_2, D_3, \dots$  closing down on the point  $O$  and such that, for each  $n$ , the boundary of  $D_n$  is compact.

<sup>4</sup> Jones, F. B., "Concerning Certain Topologically Flat Spaces," *Trans. Am. Math. Soc.*, **42**, 53-93 (1937).

<sup>5</sup> See Axiom 1.

<sup>6</sup> See the proof of Theorem 1 of Chapter I.

<sup>7</sup> See the proposition labeled "Theorem 25" in Chapter II. That this proposition is not a consequence of Axioms 0, 1 and 2 may be seen with the aid of Example 1.

## ON SETS OF INTEGERS WHICH CONTAIN NO THREE TERMS IN ARITHMETICAL PROGRESSION

BY R. SALEM AND D. C. SPENCER

DEPARTMENTS OF MATHEMATICS, HARVARD UNIVERSITY AND MASSACHUSETTS  
INSTITUTE OF TECHNOLOGY, AND STANFORD UNIVERSITY

Communicated October 26, 1942

Let  $S$  be a set of non-negative integers, all different from one another. We say briefly that  $S$  is "progression-free" if any three distinct integers of  $S$  never form an arithmetical progression, i.e., if  $a + a' \neq 2a''$  whenever  $a, a', a''$  are different and belong to  $S$ .

If the elements of a progression-free set  $S$  do not exceed a given  $N$ , then the number of elements of  $S$  has clearly a maximum  $\nu = \nu(N)$ .

It has been widely conjectured that, as  $N \rightarrow \infty$ ,  $\nu(N) = O(N)^\alpha$  where  $\alpha$  is a positive constant inferior to 1. A more precise conjecture has assigned to  $\alpha$  the value  $\log 2 / \log 3$  which corresponds to the progression-free sequence of integers whose digits in the ternary scale are 0 and 2 only.<sup>1</sup>

The purpose of the present note is to prove that the conjecture  $\nu(N) = O(N^\alpha)$  is false for every  $\alpha < 1$ . We shall prove that, as  $N \rightarrow \infty$ ,

$$\nu(N) > N^{1 - \frac{\log 2 + \epsilon}{\log \log N}}$$

for every  $\epsilon > 0$ .

Let  $d$  be an integer  $> 2$  and  $n$  an integer divisible by  $d$ . Having fixed  $d$  and  $n$ , let  $S(d, n)$  be the set of all integers given by the expression

$$A = a_1 + a_2(2d - 1) + \dots + a_n(2d - 1)^{n-1}$$

where the "digits"  $a_i$  are integers subjected to the following condition: exactly  $n/d$  digits are equal to zero,  $n/d$  digits are equal to 1,  $n/d$  digits are equal to 2, etc., and  $n/d$  digits are equal to  $d - 1$ . Thus the number of integers of  $S(d, n)$ , all different, is

$$\mu(d, n) = \frac{n!}{[(n/d)!]^d} \quad (1)$$

On the other hand, for all numbers  $A$  of  $S(d, n)$  we have

$$A < (2d - 1)^n. \quad (2)$$

The set  $S(d, n)$  is progression-free. Suppose that  $A, A', A''$  belong to the set and that  $A + A' = 2A''$ . Let  $a_i, a_i', a_i''$  be the digits of rank  $i$  in  $A, A', A''$ , respectively. Since  $a_i + a_i' \leq 2d - 2$  and  $2a_i'' \leq 2d - 2$ , the equality  $A + A' = 2A''$  implies  $a_i + a_i' = 2a_i''$  for all  $i$ . Now there are in  $A''$  exactly  $n/d$  digits equal to zero, and if  $a_h'' = 0$ , then necessarily  $a_h = a_h' = 0$ ; i.e., the  $n/d$  digits equal to zero occupy the same places in  $A, A'$  and  $A''$ . Next, there are, in  $A''$ ,  $n/d$  digits equal to 1, and if  $a_k'' = 1$ , then since  $a_k \neq 0$  and  $a_k' \neq 0$ , the equality  $a_k + a_k' = 2a_k''$  implies  $a_k = a_k' = 1$ ; i.e., the  $n/d$  digits equal to 1 correspond in  $A, A', A''$ . Generally, if  $a_i'' = m$ ,  $a_i$  and  $a_i'$  being different from  $0, 1, 2, \dots, m - 1$ , then  $a_i = a_i' = m$  and the  $n/d$  digits equal to  $m$  have the same ranks in  $A, A', A''$ . Going up to  $m = d - 1$ , we prove that  $A = A' = A''$ , that is to say  $S(d, n)$  is progression-free.

Now if  $n$  and  $n/d$  are large enough we have by (I)

$$\mu(d, n) > \frac{n^{\frac{1}{2}} \sqrt{2\pi n} e^{-n}}{[(n/d)^{n/d} \sqrt{2\pi(n/d)} e^{-n/d}]^d} \frac{1}{C^d}$$

$C$  being a constant (as near to 1 as we please). Thus

$$\mu(d, n) > (d/\gamma n)^{d/2} d^n \quad (3)$$

$\gamma$  being a constant (as near to  $2\pi$  as we please).

Let us now fix an  $N$  and let us choose  $d$  such that

$$(2d - 1)^{d\omega(d)} \leq N < (2d + 1)^{(d+1)\omega(d+1)} \quad (4)$$

where  $\omega(d)$  is an integer increasing infinitely with  $d$  and such that  $\frac{\omega(d)}{\log d} \rightarrow \infty$

but  $\frac{\log \omega(d)}{\log d} \rightarrow 0$  as  $d \rightarrow \infty$ . Let us construct the set  $S(d, n)$  with  $n = d\omega(d)$ . We have by (2), (3) and (4)

$$\begin{aligned} \nu(N) &\geq \nu[(2d - 1)^{d\omega(d)}] \geq \mu(d, d\omega(d)) > \left(\frac{1}{\gamma\omega(d)}\right)^{d/2} d^{d\omega(d)} \\ \frac{\nu(N)}{N} &> \left(\frac{1}{\gamma\omega(d)}\right)^{d/2} \frac{d^{d\omega(d)}}{(2d + 1)^{(d+1)\omega(d+1)}} \end{aligned}$$

Now, as  $N \rightarrow \infty$ ,  $d \rightarrow \infty$ , and



$$\log \left( \frac{N}{\nu(N)} \right) < (d+1)\omega(d+1) \log(2d+1) - d\omega(d) \log d +$$

$$\frac{d}{2} \log \omega(d) + \frac{d}{2} \log \gamma = d\omega(d)[\log 2 + o(1)], \quad (5)$$

if we suppose, as we may, that  $\omega(d)$  increases regularly enough to have  $\omega(d+1) - \omega(d) = o(1)$ . By (4)

$$\log N \geq d\omega(d) \log(2d-1)$$

$$\log \log N < \log(d+1) + \log \omega(d+1) + \log \log(2d+1)$$

and so

$$\frac{\log N}{\log \log N} > d\omega(d)[1 + o(1)]. \quad (6)$$

From (5) and (6) it plainly follows that, as  $N \rightarrow \infty$

$$\nu(N) > N^{1 - \frac{\log 2 + \epsilon}{\log \log N}}$$

for every  $\epsilon > 0$ .

*Remark.*—The sequence constructed above is finite and the construction depends on  $N$ . Therefore it should be pointed out that by a slight modification of the argument, we can form an infinite “progression-free” sequence of integers such that the number of terms of the sequence not exceeding  $N$  is, for  $N \rightarrow \infty$ , greater than  $N^{1 - \frac{a}{\log \log N}}$ ,  $a$  being a constant.

*Extension to Sets of Points.*—Let  $E$  be a set of points in  $(0, 1)$  such that if  $x$  and  $y$  belong to  $E$ , then  $(x+y)/2$  belongs to  $E$  if and only if  $x = y$  (property  $P$ ). It is known that  $E$  is of measure zero.<sup>2</sup> An adaptation of the above argument yields a perfect set  $E$  having the property  $P$  and whose Hausdorff dimensionality is greater than every  $\alpha < 1$ . The proof, together with other remarks on sets of points having the property  $P$ , will appear elsewhere.

<sup>1</sup> See Erdos and Turan, *Jour. London Math. Soc.*, 11, 261–264 (1936).

<sup>2</sup> See Ruziewicz, S., *Fundamenta Mathematicae* 7, 141–143 (1925).



# SOME ADVANCES IN THE COMBINATORY THEORY OF QUANTIFICATION

BY HASKELL B. CURRY

DEPARTMENT OF MATHEMATICS, THE PENNSYLVANIA STATE COLLEGE

Communicated October 17, 1942

The purpose of this paper is to report the present status of certain researches which have been interrupted by the war. These depend on a series of previous papers which will be referred to by abbreviations listed in a footnote.<sup>1</sup> The researches are those sketched in CFM, §§ 6–8. The notations of CFM and CCT will be adhered to, except as later specified.

1. *Preliminary conventions; the system  $\mathfrak{F}_0$ .* The basis of the investigation is that form of  $\lambda$ -formalism which has been proved equivalent to the system  $\mathfrak{L}_2$  (CCT, pp. 57 ff.). To this are adjoined a unary predicate, expressed as usual by prefixed ' $\vdash$ ', and an unspecified class of *canonical terms*. The rudimentary system so formed, with the rules stated below, will be called  $\mathfrak{F}_0$ .

Unspecified terms of  $\mathfrak{F}_0$  and its extensions will be denoted by capital German letters; unspecified canonical terms by small Greek letters. When the statement of a postulate, theorem, rule, etc., involves such letters it is understood that arbitrary substitutions of appropriate terms can be made. The symbols ' $\rightarrow$ ', '&', ' $\equiv$ ' will be used as in PKR, § 2.4. The symbols ' $=$ ', ' $\geq$ ' will be used, respectively, for 'is convertible to' and 'is reducible to', both in the sense of CCT, § 2.

The rules for  $\mathfrak{F}_0$  (and its extensions) are: (1) if  $\alpha = \mathfrak{B}$ , then  $\mathfrak{B}$  is canonical, and (2) the rule

RULE CONV.:  $\vdash \mathfrak{A} \ \& \ \mathfrak{A} = \mathfrak{B} \rightarrow \vdash \mathfrak{B}$ .

2. *The system  $\mathfrak{F}_1$ .* Suppose we adjoin to  $\mathfrak{F}_0$ : (1) the primitive term  $F$ , (2) the rule given for  $F$  in CFM, which will be called RULE F, (3) the rule that  $F\alpha\beta$  is always canonical, and (4) Postulates (FK) and (FS) of CFM. The system so formed will be called  $\mathfrak{F}_1$ .

In  $\mathfrak{F}_1$  all the theorems of FPF, §§ 2, 3, 4 are valid if appropriate changes (mostly replacing quantified variables by Greek letters and taking implications metatheoretically) are made. However, the proof of such theorems as 3.11 and 3.12 is facilitated by the following.

**THEOREM.** *If  $\mathfrak{X}$  is a combination of the variables  $x_1, \dots, x_n$  and the terms  $\mathfrak{A}_1, \dots, \mathfrak{A}_m$ , and if  $\alpha_1, \dots, \alpha_m$  are such that*

$$\vdash \alpha_i \mathfrak{A}_i \quad i = 1, 2, \dots, m; \quad (1)$$

*if, further, it follows from the assumptions*

$$\vdash \xi x_j \quad j = 1, 2, \dots, n$$

and the theorems (1), by following out the construction of  $\mathfrak{X}$  and applying Rule F, that

$$\vdash \eta \mathfrak{X};$$

then without hypothesis

$$\vdash F_n \xi_1 \dots \xi_n \eta (\lambda^n x_1 \dots x_n \mathfrak{X}_n).$$

This theorem can be proved directly by induction with greater ease than most of the theorems of FPF, § 3.

The consistency of  $\mathfrak{F}_1$  follows readily provided only that canonical terms are such that  $\alpha \mathfrak{X}$  can reduce only to terms of the form  $\alpha' \mathfrak{X}'$  where  $\alpha \geq \alpha'$  and  $\mathfrak{X} \geq \mathfrak{X}'$ .<sup>2</sup>

3. *The system  $\mathfrak{F}_2$ .* To form the system  $\mathfrak{F}_2$  adjoin  $\Xi$  to  $\mathfrak{F}_0$  as new primitive term and also the rule, henceforth called RULE  $\Xi$ , given for  $\Xi$  in CFM. We also need the rule that  $\Xi \alpha \beta$ ,  $\alpha \mathfrak{X}$  and  $\lambda x \alpha$ , are canonical.<sup>3</sup> Next define

$$\mathfrak{A} \supset_x \mathfrak{B} \equiv \Xi (\lambda x \mathfrak{A}) (\lambda x \mathfrak{B}).$$

The postulates are then:

$$\text{Post. } (\Xi K). \quad \vdash \alpha x \supset_x \beta x y \supset_y \alpha x$$

$$\text{Post. } (\Xi S). \quad \vdash \alpha u \supset_u \beta uv \supset_v \gamma x uv \supset_x: \alpha w \supset_w \beta w (y w) \supset_y \cdot \alpha z \supset_z \gamma x z (y z).$$

If  $F$  is defined by (10) of CFM, this system includes  $\mathfrak{F}_1$ . The following deduction theorem can also be proved:

**THEOREM.** *If from the assumption that*

$$\vdash \xi x,$$

*together with certain axioms which do not contain  $x$ , it is derivable by the rules of  $\mathfrak{F}_2$  that*

$$\vdash \mathfrak{X};$$

*then from these same axioms it is derivable absolutely within  $\mathfrak{F}_2$  that*

$$\vdash \xi x \supset_x \mathfrak{X},$$

*provided that the axioms and the terms corresponding to  $\mathfrak{A}$ ,  $\mathfrak{B}$  in every application of Rule  $\Xi$  are canonical.*

The consistency theorem of § 5 removes the last proviso. On the other hand, the theorem, apparently, cannot be iterated to take care of multiple premises such as  $\vdash \xi x$ ,  $\vdash \eta xy$ ; for when we eliminate one premise we introduce new axioms, derived from the postulates of  $\mathfrak{F}_2$ , which may contain other variables. This shows that the postulates for  $\mathfrak{F}_2$  are still in a tentative state.

4. *The system  $\mathfrak{F}_3$ .* The system obtained by adjoining to  $\mathfrak{F}_0$  the terms  $\Pi$  and  $P$  with the rules set down for them in CFM will be called  $\mathfrak{F}_3$ . For canonical terms we must have  $\Pi\alpha$ ,  $P\alpha\beta$ ,  $\alpha\tilde{x}$  and  $\lambda x\alpha$ .<sup>4</sup> The postulates will be some as yet undetermined set sufficiently strong to include  $\mathfrak{F}_2$  (with  $\Xi$  defined as in (8) of CFM) and the postulates stated in CFM.

5. *Combinatory verifiability; the systems  $\mathfrak{B}$ .* For the study of the consistency of  $\mathfrak{F}_2$  and  $\mathfrak{F}_3$  it is expedient to introduce systems  $\mathfrak{B}_2$  and  $\mathfrak{B}_3$  which have the same relation to the former that the Gentzen system LJ has to the Heyting calculus. These involve two predicates of an undetermined number of arguments, viz.,

$$\text{Can} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}) \quad n \geq 0 \quad (1)$$

$$\text{Ver} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}) \quad n \geq 0. \quad (2)$$

The interpretations of the formulas (1), (2) are " $\mathfrak{A}$  is canonical (or verifiable, respectively) on the basis  $\mathfrak{M}_1, \dots, \mathfrak{M}_n$ "; when  $n = 0$  this will mean simply that the term  $\mathfrak{A}$  is canonical or verifiable. The formula

$$\text{Reg} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}) \quad n \geq 0. \quad (3)$$

will be used to stand for either (1) or (2). A sequence of terms to which (1), (2) or (3) applies will be called a canonical, verifiable or regular sequence, respectively; in such a sequence  $\mathfrak{M}_1, \dots, \mathfrak{M}_n$  are the antecedents and  $\mathfrak{A}$  the consequent.

In the following statement of the rules ' $\left(\frac{\mathfrak{U}}{x}\right)\mathfrak{A}$ ' means the result of substituting  $\mathfrak{U}$  for  $x$  in  $\mathfrak{A}$ ; the accents should be ignored for the present.

I.  $\text{Can} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}') \rightarrow \text{Ver} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}, \mathfrak{A})$ .

II. If  $k_1, \dots, k_r$  is any sequence such that each  $k_i$  has one of the values 1, 2,  $\dots$ ,  $n$ , then

$\text{Can} (\mathfrak{M}_1, \dots, \mathfrak{M}_{n-1}, \mathfrak{M}_n') \& \text{Reg} (\mathfrak{M}_{k_1}, \dots, \mathfrak{M}_{k_r}, \mathfrak{A}) \rightarrow \text{Reg} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A})$ .

III. If  $\mathfrak{A} \geq \mathfrak{B}$ ,  $\mathfrak{M}_k \geq \mathfrak{N}_k$  then

$$\text{Reg} (\mathfrak{N}_1, \dots, \mathfrak{N}_n, \mathfrak{B}) \rightarrow \text{Reg} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}).$$

IV 1. If  $x$  does not occur in  $\mathfrak{M}_1, \dots, \mathfrak{M}_n$ , then

$$\text{Reg} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}, \mathfrak{B}) \rightarrow \text{Reg} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A} \supset \mathfrak{B}).$$

IV 2. If  $x$  does not occur in  $\mathfrak{M}_1, \dots, \mathfrak{M}_n$ , then

$$\text{Can} (\mathfrak{M}_1, \dots, \mathfrak{M}_m, \mathfrak{A}, \mathfrak{B}') \& \text{Ver} (\mathfrak{M}_1, \dots, \mathfrak{M}_m, \left(\frac{\mathfrak{U}}{x}\right)\mathfrak{A}) \& \text{Reg} (\mathfrak{M}_1, \dots, \mathfrak{M}_m,$$

$$\left(\frac{\mathfrak{U}}{x}\right)\mathfrak{B}, \mathfrak{N}_1, \dots, \mathfrak{N}_n, \mathfrak{C}) \rightarrow \text{Reg} (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A} \supset \mathfrak{B}, \mathfrak{N}_1, \dots, \mathfrak{N}_n, \mathfrak{C}).$$

These are the rules for  $\mathfrak{B}_2$ ; for  $\mathfrak{B}_3$  IV should be replaced by analogous rules for  $\Pi$  and  $P$ .

On this basis it can be shown that: (1) if a sequence is canonical, so is the subsequence obtained by omitting the consequent; (2) every verifiable sequence is canonical; (3) arbitrary substitutions can be made for the variables; (4) a theorem roughly the converse of IV holds; and (5) there is a theorem, called the *elimination theorem*, which is analogous to Gentzen's "Schnitt"—in particular it allows an antecedent to be dropped when it is verifiable on the basis of its predecessors. Also (6) canonicalness is invariant with respect to conversion.

These theorems presuppose that there are certain elementary canonical terms beginning with constants other than  $S$ ,  $K$ , for which terms there are specifications consistent with the theorems. An example of such a specification is:  $\text{Can } (Q\mathfrak{X}\mathfrak{Y})$ ,  $\text{Ver } (Q\mathfrak{X}\mathfrak{X})$ . Every term has a certain finite order, with reference to constructions by means of  $\mathfrak{E}$  from these elementary constituents; the proof of the elimination theorem involves an induction with respect to this order.

In applying this argument to the consistency of  $\mathfrak{F}_2$  and  $\mathfrak{F}_3$  certain changes in the definitions of canonicalness for these systems have to be made.<sup>6</sup> When this is done we have:

**THEOREM.** *If the axioms of the systems  $\mathfrak{F}_2$  and  $\mathfrak{F}_3$  are verifiable, then all the theorems are verifiable.*

Since  $S$ ,  $K$ ,  $\mathfrak{E}$ ,  $\Pi$ , etc., are not canonical and  $QSK$  (on the above assumption) is canonical but not verifiable, the consistency of  $\mathfrak{F}_2$  and  $\mathfrak{F}_3$  is proved.

6. *The systems  $\mathfrak{F}^*$ .* We consider now the problem of expressing the "postulates" of the systems  $\mathfrak{F}$ , which are axiom schemes, as single axioms. There are two ways of doing this: (1) by introducing quantifiers of higher order; (2) by postulating a term  $H$ , representing canonicalness, and generalizing with the quantifiers already present. The former course is not investigated here at all; the second only partially. Certain remarks concerning it are as follows.

First there must be a notion of relative canonicalness. It is not sufficient to have  $\mathfrak{E}\mathfrak{A}\mathfrak{B}$  canonical when  $\mathfrak{A}$  and  $\mathfrak{B}$  are; there must be circumstances under which  $\mathfrak{A}$  is verifiable only for certain values of  $x$ , and  $\mathfrak{B}$  canonical for the same values.

Secondly, one cannot postulate  $\vdash H(H\mathfrak{X})$ . For on the assumption as to relative canonicalness made below, there would then be, for any canonical  $\mathfrak{B}$ , a canonical term  $\mathfrak{A}$  such that  $\mathfrak{A} = H\mathfrak{A} \supset \cdot \mathfrak{A} \supset \mathfrak{B}$ , and from this a contradiction arises much as in IFL.<sup>6</sup> The system is inconsistent if  $\vdash H^n\mathfrak{X}$  is postulated for any specific  $n$ .<sup>7</sup>

Thirdly, there seems to be no use in considering  $\mathfrak{F}_2^*$ . For if  $\mathfrak{E}$  is expressible only through  $\Pi$  and  $P$ , then from  $\vdash \mathfrak{E}H\mathfrak{A}$  we must infer  $\vdash P(H\mathfrak{X})(\mathfrak{A}\mathfrak{X})$  for any  $\mathfrak{X}$ , even when neither  $H\mathfrak{X}$  nor  $\mathfrak{A}\mathfrak{X}$  is canonical. It is

difficult to see how we can get a verifiability theorem in such a case. I suspect that  $\mathfrak{F}_3^*$  (on any reasonable formulation) is inconsistent.

On the other hand, it is probably possible to formulate the system  $\mathfrak{F}_2^*$  and to prove its consistency by constructing an appropriate  $\mathfrak{B}^*$  (see § 7). Moreover, this can presumably be done so as to remove the difficulties in regard to  $\mathfrak{F}_2$  mentioned in § 3. For conjectures as to the possible significance of this for mathematics see CFM, § 8.

7. *The system  $\mathfrak{B}_2^*$ .* The formulation of the system  $\mathfrak{B}_2^*$  will not be given here in detail. However, the changes which must be made in the formulation of § 5 are essentially as follows:

1. Care must be used with variables, since now substitutions for variables cannot be made except when certain premises are fulfilled. We must distinguish three kinds of variables. A variable ' $x$ ' occurring in a specific place in a term of a sequence shall be said to be *apparent* there if that place is part of the scope of a prefix ' $\lambda x$ ', otherwise it is *real*; if it is real but is also real in some previous term of the sequence it shall be said to be *bound*; otherwise it is *free*. Substitutions are not permitted for any of these—such substitutions are taken care of by the elimination theorem. It may be desirable to have a fourth class of variables, called *indeterminates*, for which substitutions can be made. It is advisable to use different classes of letters for indeterminates and apparent variables than those which are used for real variables. Some changes to this effect should be made in IV.

2. Where the consequents of certain sequences in § 5 are primed it is to be understood that there are values for the free variables for which the formula holds. These values may be functions of the bound variables and indeterminates.

3. The suggested rules for  $H$  are

V 1.  $\text{Can } (\mathfrak{M}_1, \dots, \mathfrak{M}_n, \mathfrak{A}) \rightarrow \text{Reg } (\mathfrak{M}_1, \dots, \mathfrak{M}_n, H\mathfrak{A}).$

V 2. *If (a)  $\text{Can } (\mathfrak{M}_1, \dots, \mathfrak{M}_m, \mathfrak{A}')$*  (1)

*and if (b) from  $\text{Can } (\mathfrak{M}_1, \dots, \mathfrak{M}_m, \mathfrak{A})$*  (2)

*it is deducible<sup>8</sup> that*

$\text{Reg } (\mathfrak{N}_1, \dots, \mathfrak{N}_n, \mathfrak{B})$  (3)

*then*

$\text{Reg } (\mathfrak{M}_1, \dots, \mathfrak{M}_m, H\mathfrak{A}, \mathfrak{N}_1, \dots, \mathfrak{N}_n, \mathfrak{B}).$  (4)

4. The form of the rule V requires reformulation of the concept of a proof. If we call an ordinary proof a deduction—which is a chain of formulas connected according to the rules—we have here to do with a chain of deductions, where a whole deduction of lower order can be premise

for a formula in one of higher order. The exact formulation of this requires considerable space.

5. The induction used in the proof of the elimination theorem will be of a more complicated nature.

Can the elimination theorem be proved from these premises, or some suitable modifications of them, by a finite or transfinite induction? On that question research is unfinished.

<sup>1</sup> These abbreviations consist of the initials of the first three principal words of the title. The citations are

CCT. *Jour. Symbolic Logic*, 6, 54-61 (1941).

CFM. *Jour. Symbolic Logic*, 7, 49-64 (1942).

FPF. *Tôhoku Math. Jour.*, 41, 371-401 (1936).

IFL. *Jour. Symbolic Logic*, 7, 115-117 (1942).

PKR. *Trans. Amer. Math.* 50, 454-516 (1941).

<sup>2</sup> For it can be shown that if  $\vdash x$ , then  $x$  reduces to a term having that relation to a formula of the algebra of pure implication which was noted in CFM.

<sup>3</sup> This is an over-simplification. In  $\mathfrak{K}_{\lambda_2}$  it would be more appropriate to have different ranks of canonicalness with the rules: (1) If  $\alpha$  and  $\beta$  are canonical of rank 1, then  $\exists\alpha\beta$  is canonical of rank 0; (2) if  $\alpha$  is canonical of rank  $n$  then  $\lambda x\alpha$  is canonical of rank  $n + 1$  and, if  $n > 0$ ,  $\alpha x$  is canonical of rank  $n - 1$ . Cf. end of § 5.

<sup>4</sup> Cf. the preceding footnote.

<sup>5</sup> Cf. footnote <sup>3</sup>.

<sup>6</sup> One can, of course, introduce notions of canonicalness of higher order. For conjectures relative to this see CFM, § 8.

<sup>7</sup> Note  $H^n$  here has the significance given to the  $n$ th power by Rosser, *Annals of Mathematics*, 36, 129 (1935).

<sup>8</sup> The free variables of (2) being carried as indeterminates.



## INDEX

- Aberrations, chromosomal (Sax), 229.  
 Achlya, sexual hormones in (Raper), 509.  
 Advances in combinatory theory of quantification (Curry), 564.  
 Algal nomenclature (Papenfuss), 446.  
 Algebraic moduli (Harrison), 410.  
 Algebras, homogeneous measure (Maharam), 108.  
 Aminobenzoicless mutant (Tatum and Beadle), 234.  
 Amphibian development (Child), 339.  
 Amplitudes, photographic (Shapley), 501.  
 Analytical theory of semi-groups (Hille), 421.  
 Anaphase movement of chromosomes (Rhoades and Vilkomerson), 433.  
 ANDERS, M. V. See Sawin, P. B., 123.  
 ANDERSON, T. F. See Luria, S. E., 127.  
 Anticomplementary action, neutralization of (Tyler), 391.  
 Antifertilizin (Tyler), 391.  
 Association of three attributes (Wilson and Worcester), 384.  
 Asymptotic relations (Bateman), 371.  
 Ataxia (Sawin, Anders and Johnson), 123.  
 Atoms, metallic (Lukesh), 277.  
 Attributes, association of (Wilson and Worcester), 384.  
 Automorphisms (Miller), 368.  
 Autosomal recessive factor (Fano), 119.  
 Auxin (Wildman and Gordon), 217.  
 Avian egg (Romanoff and Bless), 306.  
*Azotobacter* (Wilson, Burris and Lind), 243.
- Bacteria, crown-gall (Braun and Laskaris), 468.  
 Bacteria, intestinal (Burkholder and McVeigh), 285.  
 Bacteriophages, identification and characterization of (Luria and Anderson), 127.  
 BARTLETT, N. R., AND HUDSON, G. E. Theory of the Effects of Light Intensity and Duration in Determining Visual Responses, 289.  
 BATMAN, H. Some Asymptotic Relations, 371.  
 ——. An Orthogonal Property of the Hypergeometric Polynomial, 374.  
 BEADLE, G. W. See Tatum, E. L., 234.  
 Bernoulli numbers (Vandiver), 324.  
 Binary systems (Kopal), 133.  
 Biochemical reactions (Tatum and Beadle), 234.  
 Bioelectric potentials of avian egg (Romanoff and Bless), 306.  
 Biorthogonal expansions (Tseng), 170.  
 BLESS, A. A. See Romanoff, A. L., 306.  
 BOAS, R. P., JR. Inversion of a Generalized Laplace Integral, 21.  
 Bodies, more than three (Garcia), 425, 428.
- BONNER, J. A Reversible Growth Inhibition of Isolated Tomato Roots, 321.  
 Boundaries (Moore), 550, 555.  
 Boundary layer problems (Weyl), 100.  
 BRAUN, A. C., AND LASKARIS, T. Tumor Formation by Attenuated Crown-Gall Bacteria in the Presence of Growth-Promoting Substances, 468.  
 BRENTON, V. K. See Payne-Gaposchkin, C., 496.  
 de BRETTEVILLE, A. See Buerger, M. J., 526.  
 BUCHANAN, J. M. See Hastings, A. B., 478.  
 BUERGER, M. J. A New Fourier Series Technique for Crystal Structure Determination, 281.  
 ——. The Characteristics of Soap Hemihydrate Crystals, 529.  
 BUERGER, M. J., SMITH, L. B., DE BRETTEVILLE, A., AND RYER, F. V. The Lower Hydrates of Soaps, 526.  
 BURKE, M. H. See Wilson, E. B., 361.  
 BURKHOLDER, P. R., AND McVEIGH, I. Synthesis of Vitamins by Intestinal Bacteria, 285.  
 ——. The Increase of B Vitamins in Germinating Seeds, 440.  
 BURRIS, R. H. See Wilson, P. W., 243.  
 B vitamins (Burkholder and McVeigh), 440.
- CAMPBELL, D. H. See Pressman, D., 77.  
 Carbon dioxide (van Niel, Ruben, Carson, Kamen and Foster), 8; (van Niel, Thomas, Ruben and Kamen), 157.  
 Carbon, radioactive (van Niel, Ruben, Carson, Kamen and Foster), 8; (van Niel, Thomas, Ruben and Kamen), 157.  
 CARSON, S. F. See van Niel, C. B., 8.  
 CASPARI, E. See Dunn, L. C., 205.  
 CASTLE, W. E. Size Genes of Mice, 69.  
 Cations, rôle of intracellular (Hastings and Buchanan), 478.  
 Cell nuclei (Mirsky and Pollister), 344.  
 Cells, division in (Sinnott), 36.  
 Cellular metabolism (van Niel, Ruben, Carson, Kamen and Foster), 8.  
 Cepheids (Shapley and McKibben), 200; (Shapley), 501.  
 Chain process (Opatowski), 83.  
 Cherry, Montmorency (Gustafson), 131.  
 CHILD, C. M. Oxidation-Reduction Patterns in Amphibian and Teleost Development, 339.  
 Chromosomal aberrations (Sax), 229.  
 Chromosome degeneration (Jones), 38.  
 Chromosomes (Wipf and Shackelford), 265; (Sutton), 268; (McClintock), 458.



- Chromosomes, anaphase movement (Rhoades and Vilkomerson), 433.  
 Chromosomes, heterobrachial (Sparrow) 463.  
 Chromosomes, X-ray and neutron effects (Marshak), 29.  
 Class of surfaces (Fubini), 178.  
 CLOOS, E. Distortion of Stratigraphic Thicknesses Due to Folding, 401.  
 Combinatory Theory of Quantification (Curry), 564.  
 Complement fixation (Pressman, Campbell and Pauling), 77.  
 Complement-release reaction (Tyler), 391.  
 Confidence intervals (Wilson), 88.  
 Congruences (Vandiver), 144, 324.  
 Contingency tables (Wilson), 94; (Wilson and Worcester), 378.  
 Continua (Moore), 544, 550.  
 Cotton oxidized (Heidelberger and Hobby), 516.  
 CRAIG, R. A. See Shapley, H., 192.  
 Crown gall-bacteria (Braun and Laskaris), 468.  
 Crystals, soap hemihydrate (Buerger), 529.  
 Crystal structure determination (Buerger), 281.  
 Cultures, toxicity of (Dubos, Hoberman and Pierce), 453.  
 CURRY, H. B. Some Advances in the Combinatory Theory of Quantification, 564.  
 Curvature, initial (Kasner and Mittleman), 48.  
 Curve, epidemic (Wilson and Burke), 361.  
 Curves, analytic (Weyl and Weyl), 417.
- DARLINGTON, C. D., AND DOBZHANSKY, T. Temperature and "Sex-Ratio" in *Drosophila Pseudoobscura*, 45.  
 DE CICCO, J. See Kasner, E., 52, 328, 413.  
 Degeneration, chromosome (Jones), 38.  
 Depolarization of neutron beams (Halpern and Holstein), 112.  
 Derivatives of a function (Schaeffer), 62.  
 Developing fruits (Sinnott), 36.  
 Diameters of galaxies (Shapley), 186.  
 Differential equations (Weyl), 100.  
 Differential equations:  $y''' = Gy'' + Hy'^2$  (Kasner), 333.  
 Differentiation, formal exponential (Vandiver), 24.  
 Diffusion of gene products (Sax), 303.  
 Dihedral groups (Miller), 368.  
 Diiodotyrosine (Johnson and Tewkesbury), 73.  
 Dirichlet problem (Hadamard), 258.  
 Dissociation constant (Wilson, Burris and Lind), 243.  
 Distortion of stratigraphic thicknesses (Cloos), 401.  
 Division, rates of cell (Sinnott), 36.  
 DOBZHANSKY, T. See Darlington, C. D., 45.
- Domains with compact boundaries (Moore), 555.  
*Drosophila Melanogaster* (Fano), 119.  
*Drosophila Pseudoobscura* (Darlington and Dobzhansky), 45.  
 Dual families (Kasner and De Cicco), 52.  
 DUBOS, R. J., HOBERMAN, H. D., AND PIERCE, C. Some Factors Affecting the Toxicity of Cultures of *Shigella dysenteriae*, 453.  
 DUNN, L. C., AND CASPARI, E. Close Linkage between Mutations with Similar Effects, 205.  
 Dynamical trajectories (Kasner and Mittleman), 48.  
 Dynamics, inverse problem of (Kasner and De Cicco), 413.
- Eclipsing binary systems (Kopal), 133.  
 Effectors, white pigmentary (Fries), 396.  
 Egg, avian (Romanoff and Bless), 306.  
 Egg, toad's (Painter and Taylor), 311.  
 EILENBERG, S., AND MACLANE, S. Natural Isomorphisms in Group Theory, 537.  
 Electron microscope (Luria and Anderson), 127.  
 Energy equation (Stewart), 161.  
 Enzymes (Wildman and Gordon), 217.  
 Epidemic curve (Wilson and Burke), 361.  
 Equations, differential (Weyl), 100; (Kasner), 333.  
 Euclidean 3-space (Kakutani), 16.  
 Expansions, generalized biorthogonal (Tseng), 170.  
 Extragalactic studies (Shapley), 186; (Shapley, McKibben and Craig), 192; (Shapley and McKibben), 200; (Shapley), 501.
- Families, isothermal and dual (Kasner and De Cicco), 52.  
 FANKHAUSER, G., AND WATSON, R. C. Heat-Induced Triploidy in the Newt, *Triturus Viridescens*, 436.  
 FANO, U. An Autosomal Recessive Factor Inducing Semi-Sterility in *Drosophila Melanogaster* Females, 119.  
 Fermat's last theorem (Vandiver), 144.  
 Fertilizin of sea-urchins (Tyler), 391.  
 Fixation, complement (Pressman, Campbell and Pauling), 77.  
 Flows, measurable (Kakutani), 16.  
 Fluid, viscous compressible (Stewart), 161.  
 Folding, stratigraphic (Cloos), 401.  
 FOSTER, J. W. See van Niel, C. B., 8.  
 Fourier series (Buerger), 281.  
 FRIES, E. F. B. White Pigmentary Effectors (*Leucophores*) in Killifishes, 396.  
 Frobenius's theorem (Miller), 251.  
 Fruitfulness in Montmorency cherry (Gustafson), 131.

- Fruits, developing (Sinnott), 36.  
 FUBINI, G. On a Class of Surfaces, 178.  
 Function, derivatives of (Schaeffer), 62.  
 Fungi, growth of (Robbins), 352.  
 Fusion, nuclear (McClintick), 458.
- Galactic studies (Shapley), 186; (Shapley, McKibben and Craig), 192; (Shapley and McKibben), 205; (Shapley), 501.  
 Galaxies, spiral (Shapley), 186.  
 GAPOSCHKIN, S. See Payne-Gaposchkin, C., 482, 490.  
 GARCIA, G. Generalization of Sundman's Fundamental Equality to the Case of More Than Three Bodies, 425.  
 ———. Generalization of the Inequality of Sundman to the Case of More Than Three Bodies and to the Case of a Gravitational Gas, 428.  
 Gas, gravitational (Garcia), 428.  
 Gene products, diffusion of (Sax), 303.  
 Genes, of mice (Castle), 69.  
 Genetic control (Tatum and Beadle), 234.  
 Genetic nature of x-ray induced changes in pollen (Rick), 518.  
 Germinating seeds (Burkholder and McVeigh), 440.  
 Glasses, two-component (Lukesh), 277.  
 Glycogen formation (Hastings and Buchanan), 478.  
 GORDON, S. A. See Wildman, S. G., 217.  
 Gravitational gas (Garcia), 428.  
 Group theory (Eilenberg and MacLane), 537.  
 Groups (Miller), 80, 141, 368, 407.  
 Groups, monothetic (Halmos and Samelson), 254.  
 Groups, transformation (Smith), 293.  
 Growth (Jones), 38.  
 Growth of fungi (Robbins), 352.  
 Growth hormone (Gustafson), 131.  
 Growth inhibition (Bonner), 321.  
 Growth-promoting substances (Braun and Laskaris), 468.  
 Growth requirements of *Spirillum Serpens* (Pennington), 272.  
 Growth substances for *phycomyces* (Robbins and Kavanagh), 4, 65.  
 Guanine (Robbins and Kavanagh), 4.  
 GUSTAFSON, F. G. Concentration of Growth Hormone and Fruitfulness in the Montmorency Cherry, 131.
- HADAMARD, J. On the Dirichlet Problem for the Hyperbolic Case, 258.  
 HALMOS, P. R., AND SAMELSON, H. On Monothetic Groups, 254.  
 HALPERN, O., AND HOLSTEIN, T. On the Depolarization of Neutron Beams by Magnetic Fields, 112.  
 Haptenic groups (Pressman, Campbell and Pauling), 77.
- HARRISON, G. The Structure of Algebraic Moduli, 410.  
 HASTINGS, A. B., AND BUCHANAN, J. M. The Role of Intracellular Cations on Liver Glycogen Formation *in Vitro*, 478.  
 Heat-induced triploidy (Fankhauser and Watson), 436.  
 HEIDELBERGER, M., AND HOBBY, G. M. Oxidized Cotton, an Immunologically Specific Polysaccharide, 516.  
 Hemihydrate crystals of soap (Buerger), 529.  
 Hereditary nervous disorder (Sawin, Anders and Johnson), 123.  
 Heterobrachial chromosomes (Sparrow), 463.  
 HILLE, E. Representation of One-Parameter Semi-Groups of Linear Transformations, 175.  
 ———. On the Analytical Theory of Semi-Groups, 421.  
 HOBBY, G. M. See Heidelberg, M., 516.  
 HOBERMAN, H. D. See Dubos, R. J., 453.  
 HOLSTEIN, T. See Halpern, O., 112.  
 Homogeneous measure algebras (Maharam), 108.  
 Hormone, growth (Gustafson), 131.  
 Hormones, sexual in *Achyla* (Raper), 509.  
 HUDSON, G. E. See Bartlett, N. R., 289.  
 Hybrid vigor (Jones), 38.  
 Hydra cluster of nebulae (Zwicky), 150.  
 Hydrates of soap (Buerger, Smith, de Bretteville and Ryer), 526.  
 Hyperbolic case (Hadamard), 258.  
 Hypergeometric polynomial (Bateman), 374.  
 Hypoxanthine (Robbins and Kavanagh), 65.
- Immunologically specific polysaccharide (Heidelberg and Hobby), 516.  
 Inactivation of tobacco mosaic virus (Marshak and Takahashi), 211.  
 Inequality of Sundman (Garcia), 428.  
 Inhibition, reversible growth (Bonner), 321.  
 Initial curvature of dynamical trajectories (Kasner and Mittleman), 48.  
 Integers, sets of (Salem and Spencer), 561.  
 Integral, Laplace (Boas), 21.  
 Intensity, light (Bartlett and Hudson), 289.  
 Intersecting continua (Moore), 544.  
 Intestinal bacteria (Burkholder and McVeigh), 285.  
 Intracellular cations (Hastings and Buchanan), 478.  
 Inverse problem (Opatowski), 83.  
 Inverse problem of dynamics (Kasner and De Cicco), 413.  
 Inversion of a generalized Laplace integral (Boas), 21.  
 Isogonal trajectories (Kasner and De Cicco), 328.

- Isomorphisms, natural (Eilenberg and MacLane), 537.  
 Isothermal families (Kasner and De Cicco), 52, 328.
- Jeannerettia (Papenfuss), 446.  
 JOHNSON, R. B. See Sawin, P. B., 123.  
 JOHNSON, T. B., AND TEWKESBURY, L. B. The Oxidation of 3,5-Dihydroxytyrosine to Thyroxine, 73.  
 JONES, D. F. Chromosome Degeneration in Relation to Growth and Hybrid Vigor, 38.
- KAKUTANI, S. Representation of Measurable Flows in Euclidean 3-Space, 16.  
 KAMEN, M. D. See van Niel, C. B., 8, 157.  
 KASNER, E. Differential Equations of the Type:  $y''' = Gy'' + Hy'^2$ , 333.  
 KASNER, E., AND DE CICCO, J. Generalized Transformation Theory of Isothermal and Dual Families, 52.  
 —. Transformation Theory of Isogonal Trajectories of Isothermal Families, 328.  
 —. Synthetic Solution of the Inverse Problem of Dynamics, 413.  
 KASNER, E., AND MITTLEMAN, D. A. General Theorem of the Initial Curvature of Dynamical Trajectories, 48.  
 KAVANAGH, F. See Robbins, W. J., 4, 65.  
 Killifishes (Fries), 396.  
 KOPAL, Z. The Calculation of Rotation Factors for Eclipsing Binary Systems, 133.
- Laplace integral (Boas), 21.  
 LASKARIS, T. See Braun, A. C., 468.  
 Leaf proteins (Wildman and Gordon), 217.  
 Leucophores (Fries), 396.  
 Light intensity (Bartlett and Hudson), 289.  
 LIND, C. J. See Wilson, P. W., 243.  
 Linear transformations (Hille), 175.  
 Linkage, close (Dunn and Caspari), 205.  
 Liver glycogen formation (Hastings and Buchanan), 478.  
 Log  $\Gamma(x)$  (Uhler), 59.  
 LUKESH, J. S. The Distribution of Metallic Atoms in Two-Component Glasses 277.  
 LURIA, S. E., AND ANDERSON, T. F. The Identification and Characterization of Bacteriophages with the Electron Microscope, 127.
- McCLINTOCK, B. The Fusion of Broken Ends of Chromosomes Following Nuclear Fusion, 458.  
 MCKIBBEN, V. See Shapley, H., 192, 200.
- MACLANE, S. See Eilenberg, S., 537.  
 McVIGOR, I. See Burkholder, P. R., 285, 440.  
 Magellanic cloud (Shapley and McKibben), 200.  
 Magnetic fields (Halpern and Holstein), 112.  
 Magnitude dispersion (Shapley, McKibben and Craig), 192.  
 MAHARAM, D. On Homogeneous Measure Algebras, 108.  
 MARSHAK, A. Relative Effects of X-Rays and Neutrons on Chromosomes in Different Parts of the "Resting Stage," 29.  
 MARSHAK, A., AND TAKAHASHI, W. N. The Effect of pH on Inactivation of Tobacco Mosaic Virus by X-Rays, 211.  
 Material ejected from novae (Payne-Gaposchkin and Gaposchkin), 482.  
 Measurable flows (Kakutani), 16.  
 Melanophores, sensitization of (Parker), 164.  
 MENGER, K. Statistical Metrics, 535.  
 Mesotrema (Papenfuss), 446.  
 Metabolism, cellular (van Niel, Ruben, Carson, Kamen and Foster), 8.  
 Metallic atoms (Lukesh), 277.  
 Metric spaces (Tseng), 170.  
 Metrics, statistical (Menger), 535.  
 Mice, size genes of (Castle), 69.  
 Microscope, electron (Luria and Anderson), 127.  
 MILLER, G. A. Maximal Sylow Subgroups of a Given Group, 80.  
 —. Certain Direct Products of the Groups of Self-Isometries, 141.  
 —. Some Deductions from Frobenius's Theorem, 251.  
 —. Automorphisms of the Dihedral Groups, 368.  
 —. The Permutation Groups of a General Degree, 407.  
 Minimal surfaces (Shiffman), 103.  
 MIRSKY, A. E., AND POLLISTER, A. W. Nucleoproteins of Cell Nuclei, 344.  
 MITTLEMAN, D. See Kasner, E., 48.  
 Moduli, algebraic (Harrison), 410.  
 Monothetic groups (Halmos and Samuelson), 254.  
 MOORE, R. L. Concerning Separability, 56.  
 —. Concerning Intersecting Continua, 544.  
 —. Concerning a Continuum and Its Boundary, 550.  
 —. Concerning Domains Whose Boundaries Are Compact, 555.  
 Montmorency cherry (Gustafson), 131.  
 Mosaic, tobacco (Marshak and Takahashi), 211.  
 Mosquitoes (Sutton), 268.

- Mutant, aminobenzoicless (Tatum and Beadle), 234.  
 Mutations (Dunn and Caspari), 205.
- Nebulae (Zwicky), 150, 317, 355.  
 Nerve cutting in melanophores (Parker), 164.  
*Neurospora* (Tatum and Beadle), 234.  
 Neutralization of anticomplementary action (Tyler), 391.  
 Neutron beams, depolarization of (Halpern and Holstein), 112.  
 Neutrons, effects on chromosomes (Marshak), 29.  
 Newt (Fankhauser and Watson), 436.  
 VAN NIEL, C. B., RUBEN, S., CARSON, S. F., KAMEN, M. D., AND FOSTER, J. W. Radioactive Carbon as an Indicator of Carbon Dioxide Utilization. VIII. The Role of Carbon Dioxide in Cellular Metabolism, 8.  
 VAN NIEL, C. B., THOMAS, J. O., RUBEN, S., AND KAMEN, M. D. Radioactive Carbon as an Indicator of Carbon Dioxide Utilization. IX. The Assimilation of Carbon Dioxide by Protozoa, 157.  
 Nitrogen fixation (Wilson, Burris and Lind), 243.  
 Nomenclature, algal (Papenfuss), 446.  
 Novae, material ejected from (Payne-Gaposchkin and Gaposchkin), 482.  
 Nuclear fusion (McClintock), 458.  
 Nuclei, cell (Mirsky and Pollister), 344.  
 Nucleic acid (Painter and Taylor), 311.  
 Nucleoproteins of cell nuclei (Mirsky and Pollister), 344.  
 Numbers, Bernoulli (Vandiver), 324.
- One-parameter semi-groups (Hille), 175.  
 OPATOWSKI, I. An Inverse Problem Concerning a Chain Process, 83.  
 Orthogonal property (Bateman), 374.  
 Oscillations of derivatives (Schaeffer), 62.  
 OSGOOD, C. Prehistoric Contact between South America and the West Indies, 1.  
 Oxidation of 3,5-diiodotyrosine to thyroxine (Johnson and Tewkesbury), 73.  
 Oxidation-reduction (Child), 339.  
 Oxidized cotton (Heidelberger and Hobby), 516.
- PAINTER, T. S., AND TAYLOR, A. N. Nucleic Acid Storage in the Toad's Egg, 311.  
 PAPERFUSS, G. F. Notes on Algal Nomenclature: *Pollexfenia*, *Jeannetia* and *Mesotrema*, 446.  
 PARKER, G. H. Sensitization of Melanophores by Nerve Cutting, 164.  
 PAYNE-GAPOSCHKIN, C., AND BRENTON, V. K. A Study of the RV Tauri Variables, 496.
- PAYNE-GAPOSCHKIN, C., AND GAPOSCHKIN, S. On the Dimensions and Constitution of Variable Stars, 490.  
 ——. On the Material Ejected from Novae, 482.  
 Patterns of oxidation-reduction (Child), 339.  
 PAULING, L. See Pressman, D., 77.  
 PENNINGTON, D. E. Purines as Growth Requirements of *Spirillum Serpens*, 272.  
 Periods for 343 Cepheids (Shapley and McKibben), 200.  
 Period-luminosity relation (Shapley, McKibben and Craig), 192.  
 Permutation groups (Miller), 407.  
 Perseus cluster of nebulae (Zwicky), 317, 355.  
 Photographic amplitudes (Shapley), 501.  
*Phycomyces* (Robbins and Kavanagh), 4, 65.  
 PIERCE, C. See Dubos, R. J., 453.  
 Pollen, x-ray induced changes in (Rick), 518.  
*Pollexfenia* (Papenfuss), 446.  
 POLLISTER, A. W. See Mirsky, A. E., 344.  
 Polynomial, hypergeometric (Bateman), 374.  
 Polysaccharide, immunologically specific (Heidelberger and Hobby), 516.  
 Potentials, bioelectric (Romanoff and Bless), 306.  
 Prehistoric contact (Osgood), 1.  
 PRESSMAN, D., CAMPBELL, D. H., AND PAULING, L. Complement Fixation with Simple Substances Containing Two or More Haptenic Groups, 77.  
 Progression, arithmetical (Salem and Spencer), 561.  
 Proteins, isolated leaf (Wildman and Gordon), 217.  
 Protozoa (van Niel, Thomas, Ruben and Kamen), 157.  
 Purines (Pennington), 272.  
 Pyrimidine analog (Robbins), 352.
- Quantification, Combinatory Theory of (Curry), 564.
- Rabbit, hereditary nervous disorder (Sawin, Anders and Johnson), 123.  
 Radioactive carbon (van Niel, Ruben, Carson, Kamen and Foster), 8; (van Niel, Thomas, Ruben and Kamen), 157.  
 RAPER, J. R. Sexual Hormones in *Achlya*. V. Hormone A', A Male-Secreted Augmenter or Activator of Hormone A, 509.  
 Recessive factor, autosomal (Fano), 119.  
 Rectifiable boundary (Shiffman), 103.  
 Red fox, chromosomes of (Wipf and Shackelford), 265.

- Reduction, oxidation (Child), 339.  
 Responses, visual (Bartlett and Hudson), 289.  
 Reversible growth inhibition (Bonner), 321.  
 RHOADES, M. M., AND VILKOMERSON, H. On the Anaphase Movement of Chromosomes, 433.  
 RICK, C. M. The Genetic Nature of X-Ray Induced Changes in Pollen, 518.  
 Rings (Vandiver), 24.  
 ROBBINS, W. J. A Pyrimidine Analog of Thiamine and the Growth of Fungi, 352.  
 ROBBINS, W. J., AND KAVANAGH, F. Guanine and Factor  $Z_1$ , Growth Substances for *Phycomyces*, 4.  
 ——. Hypoxanthine, a Growth Substance for *Phycomyces*, 65.  
 ROMANOFF, A. L., AND BLESS, A. A. The Effect of X-Ray Stimulation on the Bioelectric Potentials of the Avian Egg, 306.  
 Roots, tomato (Bonner), 321.  
 Rotation factors for binaries (Kopal), 133.  
 RV Tauri variables (Payne-Gaposchkin and Brenton), 496.  
 RUBEN, S. See van Niel, C. B., 8, 157.  
 RYER, F. V. See Buerger, M. J., 526.
- SALEM, R., AND SPENCER, D. C. On Sets of Integers Which Contain No Three Terms in Arithmetical Progression, 561.  
 Salivary gland type chromosomes (Sutton), 268.  
 SAMELSON, H. See Halmos, P. R., 254.  
 SAWIN, P. B., ANDERS, M. V., AND JOHNSON, R. B. "Ataxia," a Hereditary Nervous Disorder of the Rabbit, 123.  
 SAX, K. The Distribution of X-Ray Induced Chromosomal Aberrations, 229.  
 ——. Diffusion of Gene Products, 303.  
 SCHAEFFER, A. C. Oscillations of the Derivatives of a Function, 62.  
 Sea-urchin fertilizin (Tyler), 391.  
 Seeds, germinating (Burkholder and McVeigh), 440.  
 Self-isometries (Miller), 141.  
 Semi-groups, one-parameter (Hille), 175, 421.  
 Semi-sterility in *Drosophila Melanogaster* females (Fano), 119.  
 Sensitization of melanophores (Parker), 164.  
 Separability (Moore), 56.  
 "Sex-Ratio" in *Drosophila Pseudoobscura* (Darlington and Dobzhansky), 45.  
 Sexual hormones in *Achlya* (Raper), 509.  
 SHACKELFORD, R. M. See Wipf, L., 265.  
 SHAPLEY, H. Galactic and Extragalactic Studies, XIII. Note on the Comparative Diameters of Spheroidal and Spiral Galaxies, 186.  
 ——. Galactic and Extragalactic Studies, XVI. Photographic Amplitudes of Classical Cepheids, 501.  
 SHAPLEY, H., AND MCKIBBEN, V. Galactic and Extragalactic Studies, XV. On the Distribution of Periods for 343 Cepheids in the Small Magellanic Cloud, 200.  
 SHAPLEY, H., MCKIBBEN, V., AND CRAIG, R. A. Galactic and Extragalactic Studies, XIV. On the Magnitude Dispersion in the Period-Luminosity Relation, 192.  
 SHIFFMAN, M. Unstable Minimal Surfaces with any Rectifiable Boundary, 103.  
*Shigella dysenteriae*, cultures of (Dubos, Hoberman and Pierce), 453.  
 SINNOTT, E. W. Comparative Rates of Division in Large and Small Cells of Developing Fruits, 36.  
 Size genes of mice (Castle), 69.  
 SMITH, L. B. See Buerger, M. J., 526.  
 SMITH, P. A. Stationary Points of Transformation Groups, 293.  
 Soap (Buerger), 529; (Buerger, Smith, de Bretteville, Ryer), 526.  
 South America (Osgood), 1.  
 Spaces, unitary and metric (Tseng), 170.  
 SPARROW, A. H. Non-Random Uncoiling in Heterobrachial Chromosomes, 463.  
 SPENCER, D. C. See Salem, R., 561.  
 Spinach (Wildman and Gordon), 217.  
 Spiral galaxies (Shapley), 186.  
*Spirillum Serpens* (Pennington), 272.  
 Stars, variable (Payne-Gaposchkin and Gaposchkin), 490.  
 Stationary points (Smith), 293.  
 Statistical metrics (Menger), 535.  
 STEWART, H. J. The Energy Equation for a Viscous Compressible Fluid, 161.  
 Stirling's series for  $\log \Gamma(x)$  (Uhler), 59.  
 Storage of nucleic acid (Painter and Taylor), 311.  
 Stratigraphic thicknesses, distortion of (Cloos), 401.  
 Structure, crystal (Buerger), 281.  
 Structure of algebraic moduls (Harrison), 410.  
 Subgroups, maximal sylow (Miller), 80.  
 Substances, growth-promoting (Braun and Laskaris), 468.  
 Sundman, generalization of (Garcia), 425, 428.  
 Surfaces, class of (Fubini), 178.  
 Surfaces, unstable minimal (Shiffman), 103.  
 SUTTON, E. Salivary Gland Type Chromosomes in Mosquitoes, 268.  
 Sylow subgroups (Miller), 80.  
 Synthetic solution (Kasner and De Cicco), 413.
- TAKAHASHI, W. N. See Marshak, A., 211



- TATUM, E. L., AND BRADLE, G. W. Genetic Control of Biochemical Reactions in *Neurospora*: an "Amino-benzoicless" Mutant, 234.
- TAYLOR, A. N. See Painter, T. S., 311.
- Teleost development (Child), 339.
- Temperature in *Drosophila Pseudoöbscura* (Darlington and Dobzhansky), 45.
- TEWKESBURY, I. B. See Johnson, T. B., 73.
- Theory of Quantification (Curry), 564.
- Thiamine (Robbins), 352.
- THOMAS, J. O. See van Niel, C. B., 157.
- Tobacco mosaic (Marshak and Takahashi), 211.
- Thyroxine (Johnson and Tewkesbury), 73.
- Toad's egg (Painter and Taylor), 311.
- Tomato roots, isolated (Bonner), 321.
- Toxicity of cultures (Dubos, Hoberman and Pierce), 453.
- Trajectories, isogonal (Kasner and De Cicco), 328.
- Trajectories, dynamical (Kasner and Mittleman), 48.
- Transformation groups (Smith), 293.
- Transformation theory (Kasner and De Cicco), 52, 328.
- Transformations, linear (Hille), 175.
- Triploidy, heat-induced (Fankhauser and Watson), 436.
- Triturus Viridescens* (Fankhauser and Watson), 436.
- TSENG, Y. Y. On Generalized Biorthogonal Expansions in Metric and Unitary Spaces, 170.
- t*-Test (Wilson and Worcester), 297.
- Tumor formation (Braun and Laskaris), 468.
- Two-component glasses (Lukesh), 277.
- TYLER, A. A Complement-Release Reaction: The Neutralization of the Anticomplementary Action of Sea-Urchin Fertilizin by Antifertilizin, 391.
- UHLER, H. S. The Coefficients of Stirling's Series for  $\log \Gamma(x)$ , 59.
- Uncoiling in heterobrachial chromosomes (Sparrow), 463.
- Unitary spaces (Tseng), 170.
- VANDIVER, H. S. On Formal Exponential Differentiation in Rings, 24.
- . Certain Congruence Criteria Connected with Fermat's Last Theorem 144.
- . General Congruences Involving the Bernoulli Numbers, 324.
- Variable stars (Payne-Gaposchkin and Gaposchkin), 490.
- Variables, RV Tauri (Payne-Gaposchkin and Brenton), 496.
- Vigor, hybrid (Jones), 38.
- VILKOMERSON, H. See Rhoades, M. M., 433.
- Virus, tobacco mosaic (Marshak and Takahashi), 211.
- Viscous compressible fluid (Stewart), 161.
- Visual responses (Bartlett and Hudson), 289.
- Vitamins (Burkholder and McVeigh), 285 440.
- WATSON, R. C. See Fankhauser, G., 436.
- West Indies (Osgood), 1.
- WEYL, H. Concerning the Differential Equation of Some Boundary Layer Problems, II, 100.
- WEYL, H., AND WEYL, J. On the Theory of Analytic Curves, 417.
- WEYL, J. See Weyl, H., 417.
- White pigmentary effectors (Fries), 396.
- WILDMAN, S. G., AND GORDON, S. A. The Release of Auxin from Isolated Leaf Proteins of Spinach by Enzymes, 217.
- WILSON, E. B. On Confidence Intervals, 88.
- . On Contingency Tables, 94.
- WILSON, E. B., AND BURKE, M. H. The Epidemic Curve, 361.
- WILSON, E. B., AND WORCESTER, J. Note on the *t*-Test, 297.
- WILSON, E. B., AND WORCESTER, J. Contingency Tables, 378.
- . The Association of Three Attributes, 384.
- WILSON, P. W., BURRIS, R. H., AND LIND C. J. The Dissociation Constant in Nitrogen Fixation by *Azotobacter*, 243.
- WIPF, L., AND SHACKELFORD, R. M. Chromosomes of the Red Fox, 265.
- WORCESTER, J. See Wilson, E. B., 297, 378, 384.
- X-rays (Marshak), 29; (Marshak and Takahashi), 211; (Sax), 229; (Romanoff and Bless), 306; (Rick), 518.
- $y''' = Gy'' + Hy'''$  (Kasner), 333.
- ZWICKY, F. On the Physical Characteristics of the Hydra Cluster of Nebulae, 150.
- . On the Perseus Cluster of Nebulae, 317.
- . On the Physical Characteristics of the Perseus Cluster of Nebulae, 355.



I. A. R. I. 75.

IMPERIAL AGRICULTURAL RESEARCH  
INSTITUTE LIBRARY  
NEW DELHI.

Date of issue.	Date of issue.	Date of issue.
25 JAN 1954		
24 NOV 1954		
13.2.55		
24.7.58		
3.8.59		
29.4.64		
26 JUL 1991		
22 OCT 1992		